

**Kwaliteit van Nederlandse kinderdagverblijven:  
Trends in kwaliteit in de jaren 1995-2005**

H.J. Vermeer, M.H. van IJzendoorn, R.E.L. de Kruif, R.G. Fukkink,  
L.W.C. Tavecchio, J.M.A. Riksen-Walraven, & J. van Zeijl

Subsidiegever:

**Ministerie van Sociale Zaken en Werkgelegenheid**

Uitvoerder:

**Nederlands Consortium Kinderopvang Onderzoek**



Universiteit Leiden

Universiteit van Amsterdam

Radboud Universiteit Nijmegen

augustus 2005



## INHOUDSOPGAVE

<b>Voorwoord</b>	<b>1</b>
<b>Samenvatting</b>	<b>3</b>
<b>1. Inleiding</b>	<b>7</b>
<b>2. Methode</b>	<b>11</b>
2.1 Steekproef	11
2.2 Meetinstrumenten	13
2.3 Procedure	15
2.4 Interbeoordelaarsbetrouwbaarheid	16
<b>3. Resultaten kwaliteitsmeting 2005</b>	<b>17</b>
3.1 Structurele kenmerken	17
3.2 Proceskwaliteit	19
3.3 Relaties tussen proceskwaliteit en structurele kenmerken	25
3.4 Samenvatting kwaliteitsmeting 2005	32
<b>4. Vergelijking met kwaliteitsmetingen uit 1995 en 2001</b>	<b>35</b>
4.1 Vergelijkbaarheid	35
4.2 Proceskwaliteit	38
4.3 Relaties tussen proceskwaliteit en structurele kenmerken	43
4.4 Samenvatting vergelijking met voorgaande peilingen	43
<b>5. Conclusies en discussie</b>	<b>45</b>
<b>Referenties</b>	<b>53</b>
<b>Bijlage 1: Overzicht van de subschalen en items van de ITERS-R</b>	<b>55</b>
<b>Bijlage 2: Overzicht van de subschalen en items van de ECERS-R</b>	<b>57</b>
<b>Bijlage 3: Vergelijking van de verschillende items en subschalen</b>	<b>59</b>



---

## VOORWOORD

In dit rapport worden de resultaten van de derde landelijke peiling van de kwaliteit van groepsgebonden kinderopvang in kindercentra gepresenteerd. Eind 2004 kreeg het Nederlands Consortium Kinderopvang Onderzoek (NCKO) subsidie van het Ministerie van Sociale Zaken en Werkgelegenheid (SZW) om een landelijke peiling uit te voeren naar de kwaliteit van kinderopvang in Nederland vlak na de inwerkingtreding van de Wet kinderopvang. De onderzoeksopzet en methode zouden zoveel mogelijk moeten aansluiten bij voorgaande peilingen in Nederland.

In december 2004 werd een aanvang gemaakt met de steekproeftrekking, de werving van de kinderdagverblijven vond plaats in januari 2005 en de dataverzameling van februari tot mei 2005. Aan dit onderzoek hebben leidinggevenden en leidsters van 50 kinderdagverblijven door het gehele land hun medewerking verleend. De leidinggevenden danken wij voor het invullen van een uitgebreide vragenlijst. Naar onze mening is het een minimum vereiste voor het adequaat bepalen van de pedagogische kwaliteit van kinderopvang dat het leven in de groep gedurende minstens een ochtend geobserveerd wordt. De leidsters hebben dit mogelijk gemaakt door ons steeds, ondanks de hoge werkdruk, zeer bereidwillig in de groep toe te laten. Daarnaast hebben zij ook nog tijd vrij gemaakt voor een interview na afloop van de observaties en voor het invullen van een vragenlijst. Wij zijn zowel de leidsters als leidinggevenden, die wij hier vanwege de beloofde anonimiteit niet bij naam en toenaam kunnen noemen, zeer erkentelijk voor hun waardevolle inbreng in het onderzoek.

Wij zijn veel dank verschuldigd aan drs. E.J. Reiling. Onder haar deskundige leiding zijn de studenten intensief getraind in het afnemen van de ITERS-R. Haar bijdrage is des te waardevoller geweest voor het onderzoek, aangezien zij ook betrokken is geweest bij de trainingen ten behoeve van de kwaliteitsbepalingen van 1995 en 2001. Zo'n stabiele factor komt de vergelijkbaarheid van de drie peilingen ten goede. Tot slot willen wij de volgende studenten bedanken voor hun onontbeerlijke bijdrage aan de organisatie en dataverzameling van dit onderzoek: Annemieke Constandse, Su'en Kwok, Alev Ceylan, Canan Temur, Jourlilite Willems, Katrien Helmerhorst en Frédérique Peters. Zij hebben voor dag en dauw en door weer en wind het hele land afgereisd om in de vroege ochtenduren de observaties op de kinderdagverblijven te verrichten.

Dit project is uitgevoerd door het NCKO, bestaande uit prof. dr. L.W.C. Tavecchio, prof. dr. J.M.A. Riksen-Walraven, prof. dr. M.H. van IJendoorn, dr. H.J. Vermeer, dr. R.E.L. de Kruif, dr. R.G. Fukkink en drs. J. van Zeijl. Deze groep is gezamenlijk verantwoordelijk voor het hier gepresenteerde eindrapport.

Deze studie is financieel mede mogelijk gemaakt door het Ministerie van SZW.



---

## SAMENVATTING

In deze studie worden de resultaten beschreven van een landelijke peiling naar de kwaliteit van kinderdagverblijven anno 2005. Tevens wordt een vergelijking gemaakt met de resultaten van twee voorgaande vergelijkbare peilingen in het afgelopen decennium. Evenals in de studies van 1995 en 2001 staan de kenmerken van het zorg- en opvoedingsproces zelf, vanaf nu aangeduid met 'proceskwaliteit', centraal. Daartoe behoren onder meer de volgende aspecten: gezondheid en veiligheid, interacties met de leidsters en met de andere kinderen, en activiteiten die de ontwikkeling stimuleren. De conclusie van de peiling in 2005 luidt dat de proceskwaliteit van de kinderopvang verder is gedaald ten opzichte van de eerdere peilingen. Vooral de aspecten gezondheid, hygiëne en fysieke veiligheid blijven achter bij internationaal erkende standaards, alsook de aanwezigheid en toegankelijkheid van gevarieerd ontwikkelingsmateriaal in de kindercentra. De Nederlandse kinderopvang is haar internationale toppositie kwijtgeraakt. Deze daling is niet geconstateerd met betrekking tot de sensitiviteit van de leidster voor de kinderen, een kernaspect van het pedagogisch werk in de kinderopvang.

Het onderzoek is uitgevoerd bij een aselechte landelijke steekproef bestaande uit 50 groepen uit 50 kinderdagverblijven, verspreid over heel Nederland. Om de kwaliteit in kaart te brengen is gebruik gemaakt van twee gestandaardiseerde, internationaal veelvuldig gebruikte meetinstrumenten, namelijk de Infant Toddler Environment Rating Scale-Revised (ITERS-R; voor kinderen van 0 tot 2,5 jaar) en de Early Childhood Environment Rating Scale-Revised (ECERS-R; voor kinderen van 2,5 tot 5 jaar). Deze schalen bieden, naast een gemiddelde score voor proceskwaliteit, de mogelijkheid gemiddelde scores op het niveau van subschalen te bepalen (ruimte/meubilering, individuele zorg, taal, activiteiten, interacties, programma, ouders en staf). De kwaliteit van de leidster-kind interacties is gemeten met de Caregiver Interaction Scale (CIS). Het belangrijkste aspect van het leidstergedrag dat dit instrument meet is haar sensitiviteit in de dagelijkse omgang met de kinderen (hier aangeduid met de term 'stimulerend opvoeden'). De gegevens met betrekking tot zowel de ITERS-R/ECERS-R als de CIS zijn verkregen door middel van directe observaties gedurende een hele ochtend in de groep.

Achtergrondinformatie over de kinderdagverblijven en de leidsters (vaak aangeduid met de term 'structurele' kenmerken, bijvoorbeeld opleiding en ervaring van de leidsters en bestaansduur van het kinderdagverblijf) is verzameld aan de hand van vragenlijsten. Structurele kenmerken met betrekking tot de groep (bijvoorbeeld leidster-kind ratio) zijn verzameld tijdens de observaties in de groep.

De landelijke peiling van 2005 laat zien dat de proceskwaliteit van de kinderopvang gemiddeld op een laag niveau zit, dat wil zeggen dat er opvang wordt geboden die niet voldoet aan de internationale maatstaven die in de ITERS-R/ECERS-R zijn vastgelegd. Geen enkele van de onderzochte groepen valt, volgens de criteria van deze meetinstrumenten, wat de gemiddelde score voor proceskwaliteit betreft in de categorie 'goed'. Van de groepen valt 60% in de categorie 'middelmatic' en 40% in de categorie 'onvoldoende'. In 20 van de 50 onderzochte kinderdagverblijven wordt dus

---

opvang geboden die - volgens de normen van de ITERS-R en ECERS-R - niet voldoet aan de minimum kwaliteitseisen voor kinderopvang. Op het niveau van subschalen zijn de scores voor 'individuele zorg' het laagst (90% van de groepen scoort onvoldoende), samen met de scores voor 'activiteiten' (84% van de groepen scoort onvoldoende). De scores op de subschaal 'interacties' geven het minst aanleiding tot zorg, aangezien hier het predikaat 'goed' het vaakst is toegekend, namelijk in 36% van de groepen.

Als we inzoomen op de kwaliteit van de leidster-kind interacties zoals beoordeeld met de CIS, dan blijkt dat de leidsters in behoorlijke mate positieve 'stimulerende opvoeding' bieden en over het algemeen weinig 'autoritair' optreden. Hogere scores op de subschaal 'stimulerend opvoeden' blijken naar verwachting samen te hangen met hogere scores op de ITERS-R en ECERS-R.

Er zijn geen samenhangen gevonden tussen de proceskwaliteit van de opvang enerzijds en structurele kenmerken van de opvang anderzijds. Geen van de kenmerken op het niveau van groep, leidster en/of kinderdagverblijf blijkt significant bij te dragen aan de verklaring van verschillen in scores.

Uit de vergelijking van de peilingen in 1995, 2001 en 2005 blijkt dat de achteruitgang in de proceskwaliteit van de opvang die in 2001 werd vastgesteld, zich als een dalende trend heeft voortgezet. In vergelijking met 10 jaar geleden is er sprake van een significante kwaliteitsdaling op alle subschalen van de ITERS-R en ECERS-R. Deze daling is het meest zichtbaar voor de subschalen 'individuele zorg' en 'activiteiten' en het minst voor 'taal' en 'interacties'. Terwijl in 1995 in het geheel geen opvang van onvoldoende kwaliteit werd aangetroffen, bedroeg dit percentage in 2001 6% en was het in 2005 zelfs opgelopen tot 40% van de onderzochte groepen. In 1995 kreeg 37% van de groepen het predikaat 'goed'; in 2001 was dit percentage gedaald tot 18% en in 2005 werd dit predikaat helemaal niet meer toegekend.

Een daling in kwaliteit is niet geconstateerd als we de interacties tussen leidsters en kinderen zoals gemeten met de CIS in ogenschouw nemen. De kwaliteit van deze interacties is gemiddeld voldoende en stabiel gebleven ten opzichte van 10 jaar eerder. Deze conclusie wordt gestaafd door het gegeven dat de meeste groepen met betrekking tot de ITERS-R/ECERS-R subschalen het hoogst scoren op de subschaal 'interacties' en dat de daling in kwaliteit op deze subschaal relatief het geringst is.

In de discussiesectie van dit rapport wordt gespeculeerd over mogelijke oorzaken voor de relatief lage beoordelingen van de groepen uit de steekproef op de met de ITERS-R en ECERS-R gemeten proceskwaliteit van de opvang. Als mogelijke verklaringen worden genoemd de expansieve groei van kindercentra in het laatste decennium, met daaraan gekoppeld een tekort aan gekwalificeerd personeel en een hogere werkdruk. Tevens wordt een aantal methodologische beperkingen vermeld, die tot voorzichtigheid roepen bij het interpreteren van de vergelijkingen over de drie peilingen heen. Aanbevolen wordt in toekomstig onderzoek alle variabelen van het besproken kwaliteitsmodel voor de kinderopvang van Riksen-Walraven op te nemen, met specifieke aandacht voor de kwaliteit van de

---

dagelijkse interacties tussen leidsters en kinderen. Immers, met name deze interacties zijn tot nu toe in alle landelijke kwaliteitspeilingen ondervetegenwoordigd en slechts globaal gemeten.

Ten slotte wordt een aantal aanbevelingen gedaan voor praktijk en beleid ter verbetering van de kwaliteit van de kinderopvang in Nederland. Extra investeringen zullen zich voornamelijk moeten richten op het verbeteren van die aspecten waarop kinderdagverblijven in dit onderzoek laag beoordeeld worden, namelijk 'individuele zorg' en 'activiteiten'. Kinderdagverblijven zouden er in ieder geval zorg voor moeten dragen dat voldaan wordt aan de basale eisen van hygiëne en veiligheid. De eisen, zoals gehanteerd door de GGD, vormen hierbij een belangrijk aanknopingspunt. Verder zou er meer aandacht moeten komen voor pedagogische activiteiten in een verantwoorde omgeving. Dit vergt een materiële investering in de aanschaf van meer en gevarieerder ontwikkelingsmateriaal. Daarnaast is voor de leidsters een belangrijke taak weggelegd bij het begeleiden van activiteiten met dit materiaal. Ten slotte pleiten wij ervoor, door middel van het aanstellen van 'groepsassistenten', de zorgtaak van de leidsters te verlichten zodat zij zich voornamelijk kunnen richten op hun pedagogische taak. Deze investering lijkt ons uiterst relevant, aangezien deze niet alleen de status van het beroep verhoogt, maar ook het werkplezier en de motivatie van de leidsters ten goede komt. We verwachten dat dit uiteindelijk een positief effect zal hebben op datgene waar het werkelijk om draait binnen de kinderopvang: de ontwikkeling en het welbevinden van de kinderen.



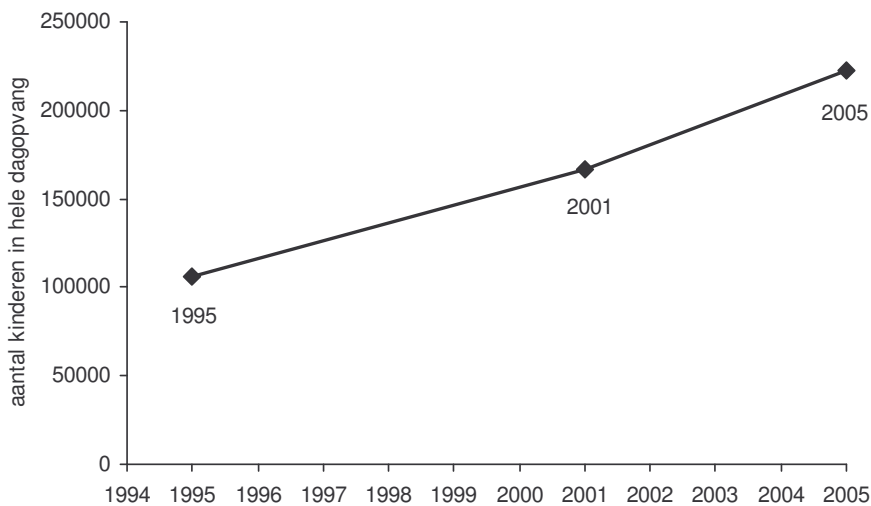
---

## 1. INLEIDING

Het doel van het hier beschreven onderzoek is het verkrijgen van een indicatie van de pedagogische kwaliteit van kinderdagverblijven in Nederland, vlak nadat de Wet kinderopvang in werking is getreden. De in dit rapport beschreven kwaliteitsmeting is de derde meting binnen hele dagopvang van het afgelopen decennium. Een vergelijking zal worden gemaakt van de kwaliteit van kinderdagverblijven anno 2005 met die van voorgaande peilingen in 1995 en 2001.

Het aantal kinderen van 0 t/m 3 jaar dat van hele dagopvang gebruik maakt is nog steeds groeiende. Ten tijde van de eerste peiling (1995) bedroeg het aantal kindplaatsen meer dan 59.000 en rond de tweede peiling (2001) was dit aantal opgelopen tot meer dan 93.000. Vlak voordat de hier gerapporteerde studie van start ging, ultimo 2004, bedroeg de capaciteit voor hele dagopvang ruim 124.000 kindplaatsen (Van der Kemp & Kloosterman, 2005). Een kindplaats voor hele dagopvang wordt gemiddeld bezet door 1.8 kind. Het aantal kinderen dat gebruik maakte van hele dagopvang ten tijde van de drie studies bedroeg dus respectievelijk rond de 106.000, 167.000 en 223.000. Anno 2005 maakt ongeveer een kwart van alle kinderen tot 4 jaar gebruik van hele dagopvang. In Figuur 1 wordt de groei van het aantal kinderen in hele dagopvang gedurende het laatste decennium grafisch weergegeven.

Figuur 1: De groei van het aantal kinderen in hele dagopvang ten tijde van de drie peilingen



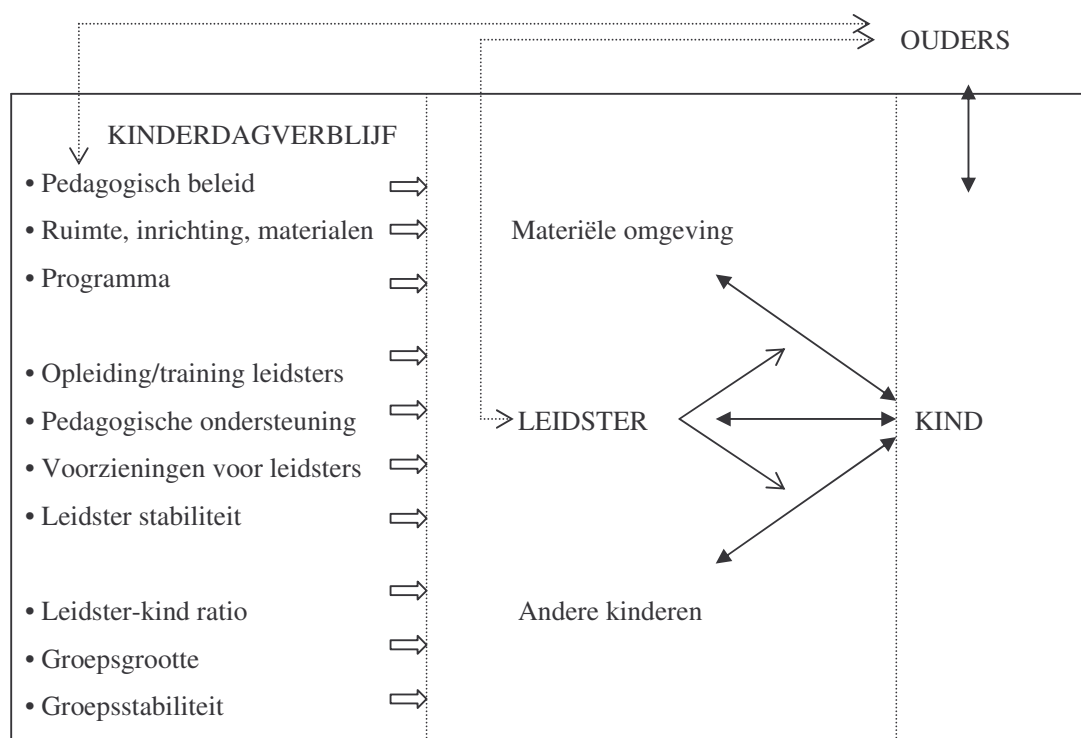
De vraag dringt zich op of de groei in kwantiteit niet ten koste is gegaan van de kwaliteit. Kwalitatief goede kinderopvang kan worden gedefinieerd als opvang die kinderen veiligheid biedt, hun persoonlijke en sociale competentie bevordert en hen regels, normen en waarden bijbrengt (Riksen-Walraven, 2004). Voor het meten van pedagogische kwaliteit<sup>1</sup> kunnen drie typen kwaliteitsindicatoren

---

<sup>1</sup> In het vervolg van deze rapportage zullen we kortweg spreken over 'kwaliteit'.

worden gebruikt: (1) maten die het welzijn en de ontwikkeling van de kinderen aangeven, (2) kenmerken van het zorg- en opvoedingsproces zelf (proceskenmerken) en (3) kenmerken van de opvoeders of het kinderdagverblijf waarvan aangenomen kan worden dat zij een goed zorg- en opvoedingsproces bevorderen (structurele kenmerken). In het model van Riksen-Walraven (zie Figuur 2) worden de onderlinge relaties tussen deze indicatoren weergegeven. In het linkerdeel van het model staan de belangrijkste structurele kenmerken waarvan is aangetoond of waarvan kan worden aangenomen dat zij van invloed zijn op de proceskwaliteit van een kinderdagverblijf. De rechterkant van het model weerspiegelt het feitelijke zorg- en opvoedingsproces (proceskwaliteit).

Figuur 2: Factoren die direct of indirect van invloed zijn op het welbevinden en de ontwikkeling van een kind in een kinderdagverblijf



Noot: ↔ Proximale processen: ontwikkeling kind in interactie met de directe omgeving  
 ⇒ Invloed van structurele opvangkenmerken op proximale processen

(Bron: Riksen-Walraven, 2004)

In de hier gerapporteerde studie hebben we niet de pretentie het gehele model op zijn geldigheid te toetsen. Op het moment van de landelijke peiling in 2005 was er (nog) geen Nederlands meetinstrument voorhanden dat alle genoemde facetten betrouwbaar in kaart brengt. Parallel aan de hier beschreven kwaliteitsmeting is het Nederlands Consortium Kinderopvang Onderzoek (NCKO) bezig met het ontwikkelen en valideren van zo'n instrument (voor projectbeschrijving, zie Tavecchio,

---

Riksen-Walraven, & Van IJzendoorn, 2002). In deze studie richten we ons, net als in de twee voorgaande peilingen, op de proceskwaliteit. Tevens onderzoeken we een aantal structurele kenmerken, die mogelijk van invloed zijn op deze proceskwaliteit. Proceskwaliteit wordt over het algemeen breed omschreven. Over welke processen in een kinderdagverblijf hebben we het eigenlijk? Volgens Cryer, Tietze, Burchinal, Leal en Palacios (1999, p. 340) omvatten deze processen "the activities that are carried out to protect children's health and safety, and to encourage their positive physical, language, intellectual, emotional, and social development". In deze omschrijving spelen, naast de leidster-kind interacties, hygiëne, de persoonlijke zorg, alsmede de typen materialen en activiteiten waarmee kinderen in aanraking komen alle een rol bij het beoordelen van de proceskwaliteit. Waar Riksen-Walraven met name de pedagogische relatie van leidsters met kinderen en de emotionele veiligheid van kinderen centraal stelt, wordt in de definitie van Cryer et al. ook het belang van de fysieke veiligheid en de materiële omgeving benadrukt. De definitie van Cryer is geen willekeurige keuze; zij is namelijk één van de medeauteurs van de meetinstrumenten die in de drie studies gebruikt zijn om een belangrijk deel van de proceskwaliteit in kaart te brengen.

Na de eerste peiling in 1995 (Van IJzendoorn, Tavecchio, Verhoeven, Reiling, & Stams, 1996; Van IJzendoorn, Tavecchio, Stams, Verhoeven, & Reiling, 1998) werd geconcludeerd dat de kwaliteit van de Nederlandse kinderopvang redelijk hoog was en dat de kindercentra de internationale vergelijking voortreffelijk konden doorstaan. Er werd een verband gevonden met de leeftijd van de leidsters: oudere leidster gaven, naar verwachting, gemiddeld kwalitatief betere opvang dan jongere leidsters. Onverwacht bleek een hogere kwaliteit ook samen te hangen met minder uren per week werken en minder ervaring in de kinderopvang. Een verklaring voor deze onverwachte resultaten werd gezocht in de zwaarte van het werk: minder uren werken en minder ervaring leiden waarschijnlijk tot minder vermoeidheid. De belangrijkste conclusie na de tweede peiling in 2001 (Gevers Deynoot-Schaub & Riksen-Walraven, 2002; Gevers Deynoot-Schaub & Riksen-Walraven, in druk) was dat de gemiddelde kwaliteit van de opvang in kindercentra was afgenomen ten opzichte van 6 jaar daarvoor. Terwijl er in 1995 in het geheel geen opvang van onvoldoende kwaliteit werd aangetroffen, was in 2001 de kwaliteit van de opvang in 6% van de onderzochte groepen onvoldoende. Opvallend was dat de verschillen in opvangkwaliteit tussen 1995 en 2001 vooral lagen op het gebied van taalervaringen en leeractiviteiten. In deze studie werd een samenhang gevonden tussen groepsgrootte en proceskwaliteit: hoe groter de groep, des te lager de opvangkwaliteit. De relatie tussen de leeftijd van de leidsters en opvangkwaliteit werd niet opnieuw aangetoond. Wel werd een samenhang gevonden met het aantal jaren ervaring op de groep: hoe meer ervaring op de groep, des te hoger de kwaliteit. Een opmerkelijke uitkomst was dat gesubsidieerde kindercentra opvang van hogere kwaliteit boden dan particuliere centra. Tevens werd gevonden dat kinderdagverblijven met een bestaansduur van 6 jaar of minder beduidend lager scoorden dan de 'oudere' kinderdagverblijven. De 'jongere' kinderdagverblijven bleken zonder uitzondering particuliere kinderdagverblijven te zijn. Het is

---

onduidelijk of het gevonden kwaliteitsverschil tussen gesubsidieerde en particuliere kindercentra voor een deel of geheel verklaard kan worden door de kortere bestaansduur van de particuliere centra.

In deze rapportage zullen de resultaten van de meting van 2005 besproken worden. Om een vergelijking met voorgaande jaren te kunnen maken, is wat betreft opzet en methode zoveel mogelijk aangesloten bij de studies van Van IJzendoorn et al. (1996) en Gevers Deynoot-Schaub en Riksen-Walraven (2002). De algemene vraagstelling luidt: Hoe is het gesteld met de kwaliteit van Nederlandse kinderdagverblijven anno 2005 en hoe verhoudt deze kwaliteit zich tot de peilingen van 1995 en 2001? Een belangrijke vraag is in hoeverre de daling in kwaliteit die in 2001 werd geconstateerd ten opzichte van de peiling in 1995 zich heeft voortgezet in de afgelopen jaren.

Na een bespreking van de steekproef, meetinstrumenten en procedure (paragraaf 2) zullen allereerst de resultaten van de peiling in 2005 afzonderlijk besproken worden (paragraaf 3). Om een beeld te schetsen van de deelnemende kinderdagverblijven wordt eerst een overzicht gegeven van de structurele kenmerken. Daarna zal worden ingegaan op de proceskwaliteit en de relaties tussen structurele kenmerken en deze twee typen van kwaliteit. In paragraaf 4 zal vervolgens een vergelijking gemaakt worden met de resultaten uit de peilingen van 1995 en 2001. Alvorens dat te doen, beargumenteren we in hoeverre de studies vergelijkbaar zijn wat betreft steekproeven, meetinstrumenten en structurele kenmerken. De resultaten zullen worden vergeleken over de drie tijdstippen heen. Ten slotte (paragraaf 5) worden de belangrijkste conclusies samengevat en bediscussieerd en worden aanbevelingen gedaan voor toekomstig onderzoek, praktijk en beleid.

---

## 2. METHODE

### 2.1 Steekproef

#### *Kinderdagverblijven*

Uit een landelijk bestand met ongeveer 2800 kinderdagverblijven (hele dagopvang), verkregen van het Programmabureau Implementatie Wet Kinderopvang, is een aselechte steekproef getrokken van 50 kinderdagverblijven verspreid over heel Nederland (steekproef 1). Het betreft een gestratificeerde steekproef, waarbij uitgegaan is van het aantal kindplaatsen per regio (prognose eind 2004). Tevens zijn twee reservebestanden gemaakt (reservesteekproef 1 en 2). We hebben gebruik gemaakt van de 21 regio's en prognoses zoals vermeld door het Netwerkbureau Uitbreiding Kinderopvang (Vervoort, Vinders, & MacLennan, 2003). Vervolgens is berekend hoeveel kinderdagverblijven per regio we naar rato nodig hebben voor de steekproef van 50.

De 50 kinderdagverblijven uit steekproef 1 hebben een brief ontvangen met informatie over het onderzoek. Vervolgens zijn deze kinderdagverblijven telefonisch benaderd voor deelname. Bij deelname werden de namen van alle groepen evenals het groepstype genoteerd, zodat een willekeurige keuze voor een groep gemaakt kon worden. Als een kinderdagverblijf niet wilde deelnemen werd de reden genoteerd en werd onmiddellijk het eerstvolgende kinderdagverblijf van de reservesteekproef uit dezelfde regio op gelijke wijze benaderd. Indien er ook een afwijzing kwam van het kinderdagverblijf uit reservesteekproef 1, werd op gelijke wijze een kinderdagverblijf van reservesteekproef 2 benaderd.

Voor steekproef 1 was er een respons van 64% (32 kinderdagverblijven). Als redenen voor het niet willen participeren werden onder meer genoemd: net nieuwe groepen gestart, te onrustig, net kwaliteitsmeting gehad, verhuizing en reorganisatie. Uiteindelijk waren 15 kinderdagverblijven uit reservesteekproef 1 en 3 kinderdagverblijven uit reserversteekproef 2 nodig om tot het beoogde aantal kinderdagverblijven te komen. In Figuur 3 wordt een overzicht gegeven van de locaties van de 50 kinderdagverblijven. Zowel de wijze waarop de steekproef is getrokken (aselect en gestratificeerd naar regio) als de omvang zorgen voor voldoende afspiegeling van het totaal aantal kinderdagverblijven in Nederland (representativiteit). Ook de typen groepen vormen een adequate afspiegeling van het totaal aantal groepen in de kinderdagverblijven.

Figuur 3: Locaties van de kinderdagverblijven



### Groepen

Van elk kinderdagverblijf werd tijdens de telefonische werving genoteerd hoeveel groepen van elk type (baby-, dreumes-, peuter- en verticale groepen) aanwezig zijn. In totaal zijn er 176 groepen op de 50 kinderdagverblijven ( $M = 3.52$ ). Opvallend is het grote aantal kinderdagverblijven met uitsluitend verticale groepen: 21 van de 50 kinderdagverblijven (42%). In Tabel 1 wordt weergegeven hoeveel groepen van elk type binnen de kinderdagverblijven respectievelijk binnen de uiteindelijke steekproef vertegenwoordigd zijn. De oorspronkelijke verhouding voor de op te nemen groepen per type was 15 babygroepen (30%), 15 peutergroepen (30%) en 20 verticale groepen (40%). Deze verhouding was gebaseerd op een schatting op basis van eerder onderzoek. Vanwege het hoge percentage verticale groepen en de verdeling van het totale aantal groepen over de kinderdagverblijven in de meting van 2005 (zie Tabel 1) is deze verhouding aangepast naar 24% baby-, 24% peuter- en 48% verticale groepen, met een toevoeging van dreumesgroepen (4%). De leeftijden van de kinderen loopt globaal van 0 tot 2 jaar voor de babygroepen, van 1 tot 2,5 jaar voor de dreumesgroepen, van 2 tot 4 jaar voor de peutergroepen en van 0 tot 4 jaar voor de verticale groepen.

Tabel 1: Verdeling van de typen groepen binnen de kinderdagverblijven en steekproef

	Groepen				Totaal
	Baby	Dreumes	Peuter	Verticaal	
Kinderdagverblijven	42 (24%)	6 (3%)	41 (23%)	87 (49%)	176 (100%)
Uiteindelijke steekproef	12 (24%)	2 (4%)	12 (24%)	24 (48%)	50 (100%)

---

Van elk kinderdagverblijf werd aselect één groep gekozen, waarbij rekening werd gehouden met het type groep. Omdat er op 21 kinderdagverblijven uitsluitend verticale groepen aanwezig waren, is het grootste deel van de verticale groepen (21 van de 24) afkomstig van deze kinderdagverblijven. Dit heeft geresulteerd in 12 baby-, 2 dreumes-, 12 peuter- en 24 verticale groepen in de steekproef. De uiteindelijke steekproef omvatte dus 50 groepen uit 50 verschillende kinderdagverblijven.

## 2.2 Meetinstrumenten

### *ITERS-R en ECERS-R*

Er is gebruik gemaakt van de Infant Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003) en de Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). Deze instrumenten zijn herziene versies van de ITERS en ECERS, zoals gebruikt in de studies van Van IJzendoorn et al. (1996) en Gevers Deynoot-Schaub en Riksen-Walraven (2002). Deze schalen zijn veelvuldig en wereldwijd gebruikt en hebben bewezen betrouwbaar en valide te zijn. De voorspellende waarde van deze schalen met betrekking tot de ontwikkeling van kinderen is herhaaldelijk aangetoond (zie bijvoorbeeld Peisner-Feinberg & Burchinal, 1997). De herziene versies zijn onlangs door het NCKO vertaald en ook in de validatiestudie (project 1; zie Tavecchio, Riksen-Walraven, & Van IJzendoorn, 2002) van het NCKO afgenomen.

Zowel de ITERS-R als de ECERS-R werden in Amerika ontwikkeld met het doel de kwaliteit van een kinderdagverblijf in brede zin te kunnen beoordelen; gesproken wordt over ‘proceskwaliteit’. De ITERS-R is bedoeld voor groepen met kinderen tot 2,5 jaar, terwijl de doelgroep van de ECERS-R bestaat uit 2,5- tot 5-jarigen. Elk van deze meetinstrumenten bestaat uit zeven subschalen, namelijk: ruimte en meubilering, individuele zorg, taal, activiteiten, interacties, programma en faciliteiten voor ouders en staf. De 39 items van de ITERS-R en 43 items van de ECERS-R worden gepresenteerd op een 7-puntsschaal met beschrijvingen (indicatoren) voor 1 (inadequaats), 3 (minimaal), 5 (goed) en 7 (uitstekend). ‘Inadequaats’ beschrijft opvang die niet voldoet aan de basale verzorgingsbehoeften, ‘minimaal’ beschrijft opvang die voldoet aan de basale ontwikkelingsbehoeften, ‘goed’ beschrijft de opvang die verder gaat en gericht is op de ontwikkeling van de kinderen en ‘uitstekend’ beschrijft opvang van hoge kwaliteit met individuele/persoonlijke zorg. Voor elk item wordt een score gegeven van 1 tot 7, waarna een gemiddelde score wordt berekend over alle items heen. Op basis van deze gemiddelde scores kunnen groepen vervolgens geclassificeerd worden in de kwaliteitsniveaus ‘onvoldoende’ (score < 3), ‘middelmatig’ (3 ≤ score < 5) en ‘goed’ (score ≥ 5) (voor een uitgebreid overzicht, zie De Kruif & Tavecchio, 2004).

### *Caregiver Interaction Scale*

De Caregiver Interaction Scale (CIS; Arnett, 1989) wordt vaak als aanvulling op de ITERS(R) of

---

ECERS(R) gebruikt om de leidster-kind interacties in kaart te brengen. Dit meetinstrument bestaat uit 26 items die worden beoordeeld op een 4-puntsschaal, variërend van 1 (helemaal niet van toepassing), 2 (enigszins van toepassing), 3 (in behoorlijke mate van toepassing) en 4 (zeer sterk van toepassing). Het belangrijkste aspect dat dit instrument meet is de sensitiviteit van de leidster in de dagelijkse omgang met de kinderen. In diverse onderzoeken worden subschalen onderscheiden om verschillende dimensies weer te geven. In de oorspronkelijke studie van Arnett (1989) worden vier dimensies onderscheiden. In sommige studies worden echter drie subschalen gebruikt (zie Tietze, Cryer, Bairrão, Palacios, & Wetzel, 1996), in andere twee of zelfs één. In de kwaliteitsmeting van 1995 (Van IJzendoorn et al., 1996) kwamen twee factoren naar voren: ‘stimulerend opvoeden’ (Cronbach’s  $\alpha$  was .90) en ‘autoritair opvoeden’ (Cronbach’s  $\alpha$  was .89). Typerende items van de eerste factor zijn: “Moedigt de kinderen aan nieuwe dingen te proberen” en “Luistert aandachtig wanneer de kinderen tegen haar praten”. Het betreft hier dus niet cognitieve ontwikkelingsstimulering in strikte zin maar een brede waaier van positief en responsief opvoedend handelen. Kenmerkende items van de tweede factor zijn: “Lijkt onnodig streng wanneer ze de kinderen een uitbrander geeft of iets verbiedt” en “Hecht grote waarde aan gehoorzaamheid”. Ook voor dit instrument geldt dat deze veelvuldig internationaal gebruikt wordt en voldoende betrouwbaar en valide is (voor een uitgebreid overzicht, zie De Kruif & Tavecchio, 2004).

#### *Vragenlijst achtergrondinformatie kinderdagverblijven*

De ‘vragenlijst achtergrondinformatie kinderdagverblijven’ bestaat uit 18 items en is gebaseerd op vragenlijsten zoals in eerder onderzoek gebruikt (Van IJzendoorn et al., 1996; De Schipper, 2003). Eén van de leidinggevendenden werd gevraagd deze vragenlijst in te vullen. De volgende onderwerpen komen aan de orde: bestaansduur, aantal kinderen, aantal kindplaatsen, aantal groepen per type, soort organisatie, flexibiliteit van de opvang, aantal leidsters en andere medewerkers, en het verloop van leidsters en andere medewerkers. Nieuwe items betreffen: (1) het gemiddeld uurtarief voor de dagopvang in de plaatsingsovereenkomst tussen het kinderdagverblijf en de ouders, (2) het al dan niet hanteren van een open-deuren beleid, en (3) het al dan niet hebben van een specifieke pedagogische visie. Omdat tijdens de telefonische werving al was gebleken dat een groot percentage kinderdagverblijven slechts verticale groepen heeft, is besloten een open vraag op te nemen naar de reden(en) voor de keuze hiervan. Een tweetal items behoeft nadere toelichting. Flexibele opvang is in de vragenlijst gedefinieerd als alle opvang die ‘afwijkt’ van de volgende definitie:

*Onder niet-flexibele opvang verstaan we opvang voor een kind in een vaste groep kinderen, met vaste leidsters, op vaste dagdelen (minimaal 4 dagdelen) tussen 8.00 ‘s morgens en 18.00 ‘s avonds in een vaste ruimte op het kinderdagverblijf (De Schipper, 2003).*

Als de vraag in hoeverre sprake is van flexibele opvang positief werd beantwoord, werd gevraagd aan te kruisen welke flexibele opvangsvormen van toepassing zijn, waarbij de keuze was uit: vroeger open, later dicht, opvang op wisselende dagen, avondopvang, nachtopvang, weekendopvang,

---

avondmaaltijden, mogelijkheid voor extra dagdelen opvang, mogelijkheid om het kind de hele dag door te brengen en te halen, en anders.

Open-deuren beleid is in de vragenlijst als volgt gedefinieerd:

*Onder open-deuren beleid verstaan we dat aan kinderen de mogelijkheid wordt geboden hun stamgroep te verlaten en gebruik te maken van verschillende ruimtes. Er moet sprake zijn van een structureel beleid; het incidenteel verlaten van de groep (om bijvoorbeeld op een centrale plek poppenkast te kijken) valt hier natuurlijk niet onder (bron: Staatscourant 17 november 2004, nr. 122, p. 18).*

Aan de leidinggevenden werd gevraagd aan te geven of er al dan niet sprake was van een open-deuren beleid.

#### *Vragenlijst achtergrondinformatie leidsters*

De 'vragenlijst achtergrondinformatie leidsters' werd ingevuld door alle vaste leidsters die op het moment van de observatie op de groep stonden. De volgende achtergrondinformatie wordt gevraagd: leeftijd, sekse, geboorteland, beroepsgerichte opleiding, hoogst genoten opleiding, aantal uren werkzaam per week, aantal jaren werkzaam binnen de kinderopvang en aantal jaren werkzaam binnen de betreffende groep.

### **2.3 Procedure**

De ITERS-R, ECERS-R en CIS werden afgenomen door derde- en vierdejaars studenten van de universiteiten van Amsterdam en Leiden. Deze studenten hadden een intensieve training ondergaan op de betreffende instrumenten in het kader van de validatiestudie van het NCKO (project 1). Zij waren betrouwbaar gebleken op deze instrumenten volgens de gebruikelijke normen van de auteurs en hadden ruime ervaring opgedaan met het afnemen van deze instrumenten tijdens de validatiestudie. Twee studenten waren getraind op de ITERS-R en vier studenten waren getraind op de ECERS-R. Alle zes studenten waren bovendien betrouwbaar gebleken op het afnemen van de CIS.

Alle observaties vonden plaats in de periode februari tot en met april 2005. De ITERS-R werd in alle babygroepen en dreumesgroepen afgenomen. Omdat in één verticale groep de meeste kinderen onder de 2,5 jaar oud waren, werd ook hier de ITERS-R afgenomen. Dit resulteerde in 15 ITERS-R afnames. In de overige groepen ( $N = 35$ ) werd een ECERS-R afgenomen. Elke observator verbleef een hele ochtend in de betreffende groep. Aan het eind van de ochtend werd voor elke vaste leidster die op dat moment in de groep verbleef een CIS ingevuld. Met één van de leidsters vond een aanvullend interview plaats, om informatie te verkrijgen over items van de ITERS-R respectievelijk ECERS-R, die niet via directe observaties te verkrijgen was.

De vragenlijsten voor de leidinggevenden en leidsters werden, direct na de observatie, per post opgestuurd. Er werd voor deze procedure gekozen, omdat we dan de namen van de leidsters, bij wie

---

een CIS was afgenomen, tot onze beschikking hadden. Via deze weg konden we de leidsters persoonlijk benaderen een vragenlijst in te vullen. Na een aantal weken werden de kinderdagverblijven van wie we nog geen vragenlijsten retour hadden ontvangen, telefonisch benaderd. Dit leidde tot een respons van 100% voor de leidinggevend en 98% voor de leidsters (49 leidsters). Omdat er in de vragenlijst voor de leidinggevend een vraag was opgenomen of het kinderdagverblijf al of niet gesubsidieerd werd en dit niet meer van toepassing is met de invoering van de nieuwe Wet kinderopvang, is besloten alle 50 kinderdagverblijven telefonisch te benaderen om te vragen naar de situatie vóór 1 januari 2005. Tevens werd van deze mogelijkheid gebruik gemaakt om na te gaan in hoeverre het kinderdagverblijf in het bezit is van een Harmonisatie Kwaliteitsbeoordeling in de Zorgsector (HKZ) certificering. Van alle 50 kinderdagverblijven is deze informatie verkregen, zowel met betrekking tot de subsidiestatus als met betrekking tot de HKZ certificering.

## **2.4 Interbeoordelaarsbetrouwbaarheid**

### *ITERS-R/ECERS-R*

Om na te gaan in hoeverre de observatoren - ruim na hun training - nog betrouwbaar coderen, is vijfmaal tijdens de observaties (10%) 'dubbel' gecodeerd, dat wil zeggen dat de observatoren in tweetallen observeerden en onafhankelijk van elkaar een beoordeling gaven. Dit is eenmaal gebeurd voor de ITERS-R en viermaal voor de ECERS-R. Deze verhouding is gekozen, omdat er minder groepen zijn waarbinnen een ITERS-R ( $n = 15$ , namelijk 12 babygroepen, 2 dreumesgroepen en 1 verticale groep) is afgenomen dan een ECERS-R ( $n = 35$ , namelijk 12 peutergroepen en 23 verticale groepen). Ook waren er slechts twee ITERS-R observatoren, terwijl er vier studenten getraind waren op de ECERS-R (zie paragraaf 2.3). De interbeoordelaarsbetrouwbaarheid werd op dezelfde manier bepaald als binnen de training, namelijk door het percentage overeenstemming in scoring tussen beide observatoren te berekenen binnen een schaalpunt. De gemiddelde betrouwbaarheid was 86%, hetgeen ruim voldoende is.

### *CIS*

Ook voor de CIS geldt, dat er tijdens vijf observaties dubbel gecodeerd is. De interbeoordelaarsbetrouwbaarheid is op dezelfde manier berekend als tijdens de training, namelijk het percentage overeenstemming in scoring met maximaal een schaalpunt verschil, behalve tussen de schaalpunten 2 en 3. Dit is op deze manier gebeurd omdat een verschil tussen de schaalpunten 1 en 2 ('helemaal niet' en 'enigszins') en tussen 3 en 4 ('in behoorlijke mate' en 'zeer sterk') als minder kwalijk wordt gezien dan een verschil tussen de schaalpunten 2 en 3. De gemiddelde betrouwbaarheid was voldoende, namelijk een overeenstemming van 73%.

### 3. RESULTATEN KWALITEITSMETING 2005

#### 3.1 Structurele kenmerken

In Tabel 2 wordt een overzicht gegeven van de structurele kenmerken van de kinderdagverblijven uit de steekproef op groeps-, leidster- en kinderdagverblijfniveau.

Tabel 2: Structurele kenmerken

	<i>M</i>	<i>SD</i>	Range
<i>Groep (N = 50)</i>			
Groepsgrootte	9.7	2.6	5 – 17
Leidster-kind ratio	.22	.06	.10 - .33
<i>Leidsters (N = 49)</i>			
Leeftijd	29.2	7.1	20 – 51
Ervaringsjaren in kinderopvang	6.4	4.5	1 – 24
Ervaringsjaren in groep	3.2	2.1	.75 – 9
Uren aanstelling	30.0	6.1	10 – 40
<i>Kinderdagverblijven (N = 50)</i>			
Bestaansduur (in jaren)	10.0	6.8	2 – 31
Grootte (aantal kindplaatsen)	52.7	28.9	6 – 145
Uurtarief (€)	5.2	0.5	3.5 - 6.1
	<i>N (%)</i>		
Slechts verticale groepen <sup>a</sup>	21 (42%)		
Specifieke pedagogische visie <sup>a</sup>	12 (24%)		
Gesubsidieerd <sup>a</sup>	25 (50%)		
Flexibel <sup>a</sup>	34 (69%)		
Open deurenbeleid <sup>a</sup>	23 (46%)		
HKZ-certificaat <sup>a</sup>	13 (26%)		

*Noot.* <sup>a</sup>Deze variabelen zijn dichotoom gescoord: wel/niet van toepassing.

#### *Groepen*

De informatie over de groepen zoals weergegeven in Tabel 2 is verkregen aan de hand van directe observaties in de groepen, namelijk tijdens de afname van de ITERS-R of ECERS-R. Tijdens de observaties waren er gemiddeld bijna 10 kinderen aanwezig in de groep, en was er gemiddeld 1 leidster op ongeveer 5 kinderen (ratio = .22). Op de babygroepen waren er gemiddeld 8.7 kinderen in de groep (4.2 kinderen per leidster), op de peutergroepen gemiddeld 12.3 kinderen (6.1 kinderen per leidster) en op de verticale groepen gemiddeld 9 kinderen (4.8 kinderen per leidster).

---

### *Leidsters*

Van 49 leidsters hebben we een vragenlijst teruggekregen. Deze leidsters zijn allen van het vrouwelijk geslacht; 45 leidsters (92%) zijn in Nederland geboren. Eenendertig leidsters (62%) hebben een opleiding op MBO niveau, 12 leidsters (25%) hebben een afgeronde opleiding onder MBO-niveau en 6 leidsters (12%) op HBO of WO-niveau. In Tabel 2 wordt een samenvatting gegeven van de overige kenmerken. De leidsters zijn gemiddeld 29 jaar, hebben gemiddeld 6,4 jaar ervaring in de kinderopvang, waarvan iets meer dan 3 jaar op de betreffende groep en werken gemiddeld 30 uur per week.

### *Kinderdagverblijven*

De kinderdagverblijven uit de steekproef hebben een gemiddelde bestaansduur van 10 jaar (zie Tabel 2). Hierin zit echter een grote spreiding: het 'jongste' kinderdagverblijf bestaat 2 jaar en het 'oudste' 31 jaar. Het aantal kindplaatsen varieert van 6 tot 145, met een gemiddelde van bijna 53. Er zijn geen noemenswaardige verschillen in de gemiddelde uurtarieven die de kinderdagverblijven hanteren (rond de 5 euro).

*Verticale groepen.* Indien een kinderdagverblijf aangaf alleen uit verticale groepen te bestaan, werd gevraagd sinds wanneer deze situatie bestaat en naar de reden(en) hiervoor (open vraag). Deze vraag werd door alle 21 kinderdagverblijven met alleen verticale groepen beantwoord. De gemiddelde tijdsduur blijkt bijna 6 jaar te zijn. Bij vier kinderdagverblijven bestaat deze situatie korter dan een jaar; bij zeven kinderdagverblijven tussen de 2 en 4 jaar; bij drie kinderdagverblijven tussen de 4 en 6 jaar, bij vier kinderdagverblijven tussen de 6 en 8 jaar en bij drie kinderdagverblijven langer dan 8 jaar, inclusief een uitschieter van 31 jaar. Als redenen voor de keuze van uitsluitend verticale groepen gaf het grootste deel van de 21 kinderdagverblijven ( $n = 13$ ; 62%) alleen pedagogische redenen aan; 2 kinderdagverblijven vermeldden slechts organisatorische redenen, en 5 kinderdagverblijven gaven zowel pedagogische als organisatorische redenen aan. In één kinderdagverblijf werd gerapporteerd dat dit een beleid van de overkoepelende organisatie was. In totaal werden pedagogische redenen door leidinggevendenden van 18 kinderdagverblijven aangegeven. De volgende redenen werden het vaakst genoemd: (1) nabootsen van de gezinssituatie (10 keer), (2) kinderen leren van en stimuleren elkaar (10 keer), (3) continuïteit voor kinderen (geen wisselende leidsters, opbouwen van band met leidsters) (6 keer). Opgemerkt moet worden, dat het totaal groter is dan 18, aangezien een leidinggevende meerdere redenen kon aangeven. Als organisatorische redenen werden onder andere genoemd dat de plaatsing gemakkelijker verloopt in verticale groepen en dat het betreffende kinderdagverblijf zo weinig groepen heeft, dat het vormen van verticale groepen de enige mogelijkheid is.

*Specifieke pedagogische visie.* Ongeveer een kwart van de kinderdagverblijven vermeldt een specifieke pedagogische visie aan te hangen. De volgende pedagogische stromingen worden genoemd: Gordon (4 keer), Reggio Emilia (3 keer), Steiner (1 keer), combinatie van Gordon en Reggio (1 keer), combinatie van Gordon en Korczak (1 keer), een combinatie van Steiner, Freinet, Malaguzzi en

---

Gordon (1 keer) en ontwikkelingsgericht werken (1 keer).

*Gesubsidieerd.* De helft van de kinderdagverblijven gaf in het telefonische interview aan (deels) gesubsidieerde plaatsen te hebben in de periode vóór 1 januari 2005.

*Flexibel.* Als een kinderdagverblijf had aangegeven flexibele opvang aan te bieden, werd gevraagd naar specifieke flexibele opvangsvormen die al dan niet aanwezig kunnen zijn. De mogelijkheid van avond-, nacht-, en/of weekendopvang werd geen enkele keer genoemd door de kinderdagverblijven ( $n = 34$ ) die aangaven flexibele opvang te bieden. De volgende flexibele opvangsvormen werden aangekruist: (1) extra dagdelen (27 keer; 79%), (2) vroeger open (25 keer; 74%), (3) wisselende dagen (22 keer; 65%), (4) hele dag brengen/halen (11 keer; 32%), (5) later dicht (18 keer; 53%), (6) avondmaaltijd (7 keer; 21%). Alle 34 kinderdagverblijven hadden minstens 2 flexibele opvangsvormen aangekruist.

*Open deurenbeleid.* Bijna de helft van de kinderdagverblijven (48%) gaf aan een open deurenbeleid te hanteren.

*HKZ-certificaat.* Dertien van de 50 kinderdagverblijven (26%) is in het bezit van een HKZ-certificaat. Van de overige 37 kinderdagverblijven zijn er 5 bezig met het verkrijgen van zo'n certificaat; 3 zijn er onbekend mee en één kinderdagverblijf is hier een tegenstander van.

### 3.2 Proceskwaliteit

#### *ITERS-R en ECERS-R*

In Tabel 3 wordt een overzicht gegeven van de gemiddelde scores en standaarddeviaties van de groepen op de ITERS-R, respectievelijk ECERS-R.

Tabel 3: Beschrijvende statistiek voor de ITERS-R en ECERS-R

	ITERS-R ( $N = 15$ )			ECERS-R ( $N = 35$ )		
	<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	Range
Totaal	3.4	0.46	2.9 - 4.6	3.0	0.52	2.1 - 4.1
<i>Subschalen</i>						
Ruimte/meubilering	3.2	0.50	2.4 - 4.4	3.5	0.71	1.9 - 5.1
Individuele zorg	2.4	0.31	1.7 - 2.8	2.4	0.91	1.5 - 5.8
Taal	3.7	0.92	2.7 - 5.7	3.3	1.09	1.0 - 5.2
Activiteiten	3.0	0.68	2.4 - 4.7	2.1	0.44	1.3 - 3.1
Interacties	5.0	1.12	3.3 - 6.8	3.9	1.33	1.0 - 6.0
Programma	4.7	0.95	3.3 - 6.7	3.5	0.76	2.0 - 5.0
Ouders en staf <sup>a</sup>	4.9	0.46	4.0 - 5.6	4.3	1.06	2.4 - 6.7

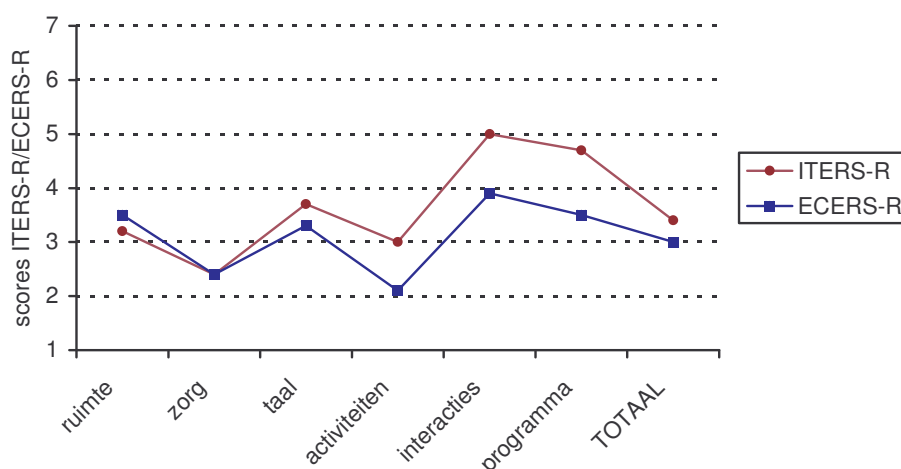
<sup>a</sup>Deze subschaal is niet opgenomen in de totaalscore

Evenals in voorgaande Nederlandse studies en internationale studies (zie bijvoorbeeld Cryer et al., 2005), is bij het bepalen van de gemiddelde scores - zowel bij de ITERS-R als de ECERS-R - de subschaal 'ouders en staf' weggelaten, omdat de items uit deze subschaal niet de directe dagelijkse ervaringen van de kinderen reflecteren. De interne consistentie (Cronbach's  $\alpha$ ) van de schalen was .83 voor de ITERS-R en .80 voor de ECERS-R. Volgens de normen van de ITERS-R en ECERS-R is de gemiddelde score van de ECERS-R 'minimaal' ( $M = 3.0$ ), en ligt die van de ITERS-R iets boven 'minimaal' ( $M = 3.4$ ).

Voor zowel de ITERS-R als ECERS-R geldt, dat de laagste scores worden gehaald voor de subschalen 'individuele zorg' (beide  $M = 2.4$ ) en 'activiteiten' (respectievelijk  $M = 3.0$  en  $M = 2.1$ ). De hoogste scores zijn gegeven op de subschaal 'interacties' ( $M = 5.0$  respectievelijk  $M = 3.9$ ); ook de spreiding in scores is hier echter het grootst. Voor zowel de ITERS-R als de ECERS-R wordt een relatief hoge score gehaald op de subschaal 'ouders en staf', maar deze is, net als in voorgaande jaren, niet opgenomen in de totaalscore. Met betrekking tot de ECERS-R totaalscores blijkt er geen verschil te zijn in scores tussen peutergroepen ( $M = 3.0$ ;  $SD = .60$ ) en verticale groepen ( $M = 3.0$ ;  $SD = .49$ ).

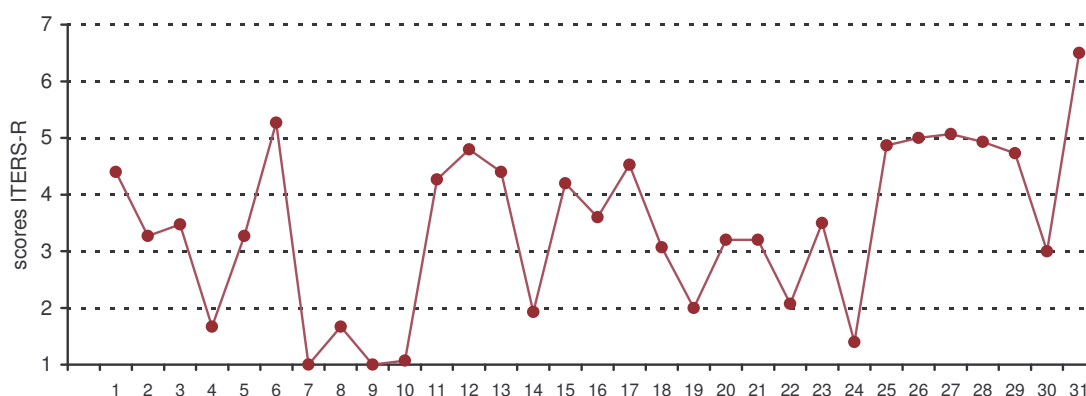
In Figuur 4 worden de gemiddelde scores voor de totaalscores en subschalen voor de ITERS-R en ECERS-R afzonderlijk grafisch weergegeven. De figuur illustreert de relatieve piek voor 'interacties' en het relatieve dal voor 'individuele zorg' en 'activiteiten' voor beide schalen.

Figuur 4: Gemiddelde scores op de ITERS-R/ECERS-R



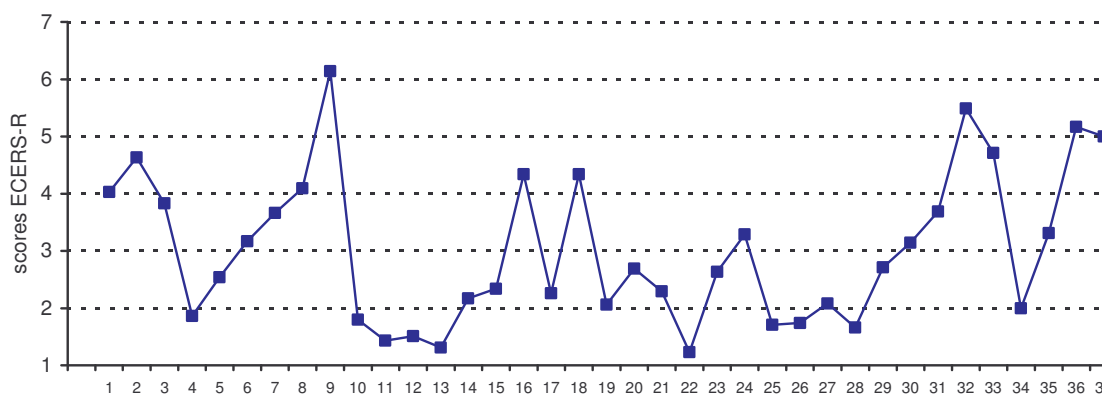
Het is tevens interessant na te gaan in hoeverre er voor de ITERS-R en ECERS-R uitschieters zijn op itemniveau, hetzij naar boven hetzij naar beneden. Zijn er items waarop consequent laag respectievelijk hoog wordt gescoord en wat is de inhoud van deze items? In de bijlagen 1 en 2 wordt een overzicht gegeven van de subschalen en items van respectievelijk de ITERS-R en ECERS-R. In Figuur 5 en 6 worden de gemiddelde scores op itemniveau weergegeven.

Figuur 5: Vergelijking van gemiddelde ITERS-R scores op itemniveau



Figuur 5 laat zien dat voor alle groepen twee items (item 7 en item 9) van de ITERS-R consequent de laagste beoordeling krijgen (score 1) en dat dit voor één item (item 10) voor bijna alle groepen geldt ( $M = 1.1$ ). Item 7 (zie bijlage 1) verwijst naar ‘maaltijden en tussendoortjes’; de groepen komen niet verder dan een inadequate beoordeling omdat de basale hygiënische procedures gewoonlijk worden verzuimd. Voor de items 9 (‘verschonen en toiletgebruik’) en 10 (‘gezondheidspraktijk’) geldt hetzelfde. Een uitschieter naar boven is duidelijk item 31 ( $M = 6.5$ ) en, in mindere mate, item 6 ( $M = 5.3$ ). Item 31 verwijst naar ‘spelactiviteiten in een groep’. Een hoge score hierop wordt behaald als de grootte van de groep geschikt is voor de leeftijden van de kinderen en de activiteiten, de staf flexibel is en alternatieve activiteiten toegankelijk maakt voor kinderen die niet meedoen en tegemoet komt aan de individuele behoeften van kinderen. Item 6 omvat het brengen en halen. Een hoge score op dit item wordt behaald als de staf sensitief is - ook bij problemen bij het afscheid nemen -, er een vriendelijke en ontspannen sfeer heerst en er een individueel, schriftelijk verslag over het verloop van de dag aan de ouders wordt aangeboden.

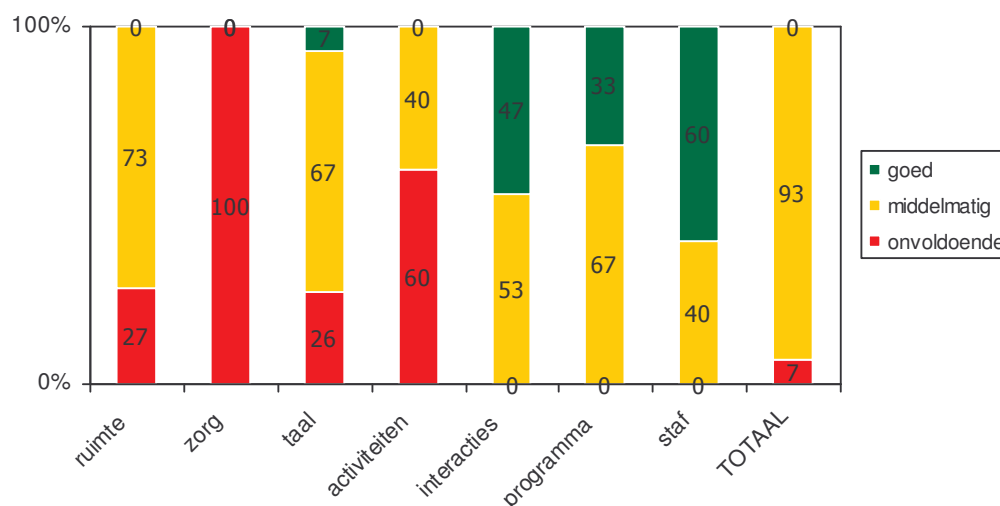
Figuur 6: Vergelijking van gemiddelde ECERS-R scores op itemniveau



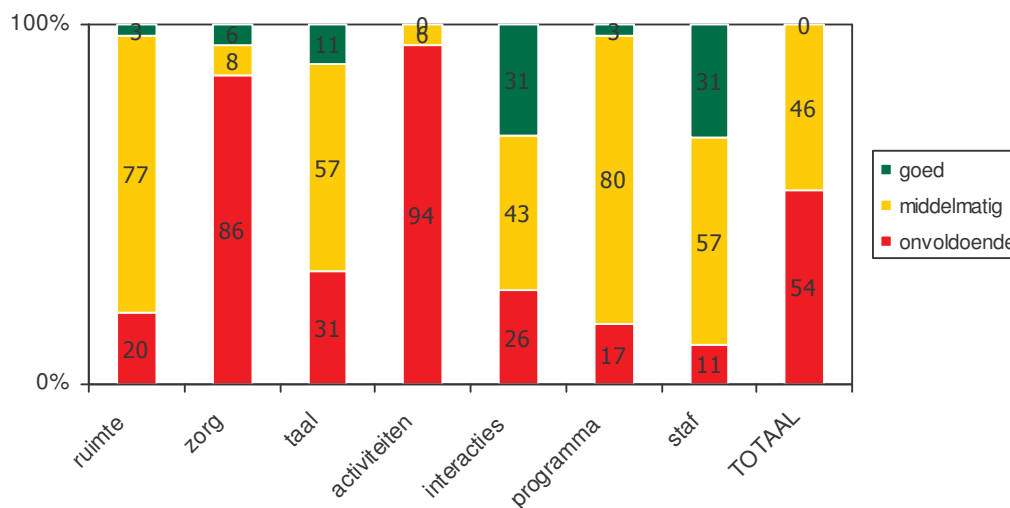
Uit Figuur 6 blijkt dat zeer lage scores op de ECERS-R vooral voorkomen bij de items 13 ( $M = 1.3$ ) en 22 ( $M = 1.2$ ). Item 13 komt overeen met item 10 van de ITERS-R ('gezondheidspraktijk', zie vorige bladzijde). Op item 22 ('blokken') wordt laag gescoord, omdat er simpelweg op de meeste kinderdagverblijven geen of niet genoeg blokken aanwezig zijn en/of niet toegankelijk voor de kinderen. Relatief hoge scores worden behaald voor item 9 ( $M = 6.1$ ) en item 32 ( $M = 5.5$ ). Item 9 is gelijk aan item 6 van de ITERS-R; ook voor het brengen en halen in peutergroepen en verticale groepen worden hoge scores gegeven. Item 32 ten slotte heeft betrekking op de leidster-kind interacties. Voor het behalen van een hoge score moeten de leidsters warmte en respect tonen, sympathiek reageren als kinderen aangeven pijn te hebben of boos te zijn en uitstralen dat ze het leuk vinden met de kinderen op te trekken.

Vervolgens is nagegaan hoe de groepen verdeeld zijn - zowel voor ITERS-R als ECERS-R scores - over de kwaliteitsniveaus 'onvoldoende' (score < 3), 'middelmatig' ( $3 \leq \text{score} < 5$ ) en 'goed' (score  $\geq 5$ ). De resultaten zijn weergegeven in Figuur 7 en Figuur 8. Uit Figuur 7 en 8 blijkt dat voor de totale score van de ITERS-R geldt, dat één groep 'onvoldoende' scoort (7%), terwijl voor de ECERS-R 19 groepen (54%) in de categorie 'onvoldoende' vallen. Geen enkele groep valt in de categorie 'goed'. Op subschaal-niveau kunnen we verder het volgende concluderen: Voor zowel de ITERS-R als de ECERS-R geldt dat de categorie 'goed' het meest voorkomt bij de subschaal 'interacties'. Opgemerkt moet worden dat hierbij de subschaal 'ouders en staf' buiten beschouwing is gelaten, omdat deze niet is opgenomen in de totaalscore. Wat betreft de ITERS-R vallen 7 groepen (47%) in deze categorie en wat betreft de ECERS-R 11 groepen (31%). Verder valt op dat alle groepen, waarbij een ITERS-R is afgenomen, onvoldoende scores op de subschaal 'zorg'; voor de ECERS-R geldt dit voor 30 kinderdagverblijven (86%). Op de subschaal 'activiteiten' wordt vooral laag gescoord bij gebruik van de ECERS-R: 94% van de groepen valt in de categorie 'onvoldoende'. Bij gebruik van de ITERS-R valt 60% in de categorie 'onvoldoende' voor deze subschaal.

Figuur 7: Verdeling van de ITERS-R scores ( $N = 15$ ) over de categorieën goed, middelmatig en onvoldoende

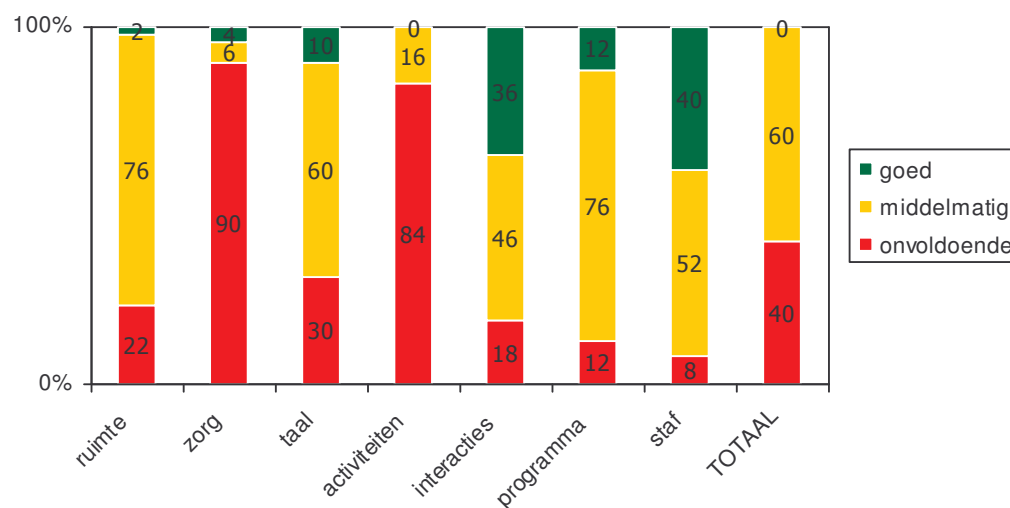


Figuur 8: Verdeling van de ECERS-R scores ( $N = 35$ ) over de categorieën goed, middelmatig en onvoldoende



In Figuur 9 wordt een totaaloverzicht gegeven van de verdeling van scores van de gehele steekproef, waarbij de scores van de ITERS-R en ECERS-R zijn samengevoegd. Uit deze figuur blijkt dat 60% (30 groepen) het predikaat ‘middelmatig’ krijgt en 40% (20 groepen) het predikaat ‘onvoldoende’. Op het niveau van de subschalen worden de meeste onvoldoendes gegeven voor ‘zorg’ (90%) en ‘activiteiten’ (84%). De categorie ‘goed’ wordt het vaakst toegekend bij de subschalen ‘interacties’ (36%) en ‘ouders en staf’ (40%).

Figuur 9: Verdeling van de gecombineerde ITERS-R/ECERS-R scores ( $N = 50$ ) over de categorieën goed, middelmatig en onvoldoende



---

## CIS

Voor alle leidsters die op het moment van een ITERS-R of ECERS-R observatie op een groep waren (gemiddeld twee), is een CIS ingevuld. Om de analyses zuiver te houden – er is afhankelijkheid tussen de scores van beide leidsters en de scores op de ITERS-R en ECERS-R, die immers op groepsniveau worden gemeten - hebben we van elke groep slechts één leidster in de analyses betrokken. Van elke groep is aselekt een leidster gekozen ( $N = 50$ ). Vervolgens is op de 26 items van de CIS een Principale Componenten Analyse uitgevoerd. Net als in de studie van Van IJzendoorn et al. (1996) werden twee factoren gevonden, namelijk ‘stimulerend opvoeden’ (14 items; Cronbach’s  $\alpha = .93$ ) en ‘autoritair opvoeden’ (12 items; Cronbach’s  $\alpha = .80$ ). De factor ‘stimulerend opvoeden’ bevat items die refereren aan positieve aandacht en sensitiviteit van de leidster (bijv. "Luistert aandachtig wanneer de kinderen tegen haar praten"), en moet dus niet verward worden met cognitieve ontwikkelingsstimulering. De andere factor, ‘autoritair opvoeden’ bevat items die onder andere verwijzen naar strengheid, discipline en gehoorzaamheid (bijv. "Hecht grote waarde aan gehoorzaamheid"). ‘Stimulerend opvoeden’ is de positieve dimensie van deze schaal; een hoge score op deze subschaal geeft aan dat leidsters over het algemeen warmte tonen, positieve aandacht geven aan de kinderen en plezier lijken te beleven aan de kinderen. ‘Autoritair opvoeden’ is de negatieve dimensie van deze schaal; een hoge score op deze subschaal impliceert juist dat leidsters over het algemeen grote waarde hechten aan gehoorzaamheid, proberen sterke controle op de kinderen uit te oefenen en verwachten dat de kinderen zelfbeheersing vertonen. De correlatie tussen deze twee factoren is  $-.16$ . In Tabel 4 wordt een overzicht gegeven van de gemiddelde scores van de leidsters voor de verschillende typen groepen.

Tabel 4: Beschrijvende statistiek voor de scores op de CIS

	<i>N</i>	Stimulerend			Autoritair		
		<i>M</i>	<i>SD</i>	range	<i>M</i>	<i>SD</i>	range
Totaal	50	3.0	0.57	1.9 - 4.0	1.8	0.46	1.1 - 3.3
Baby <sup>1</sup>	14	3.1	0.41	2.0 - 3.8	1.5 <sub>a</sub>	0.25	1.3 - 1.9
Peuter	12	2.7	0.60	1.9 - 3.8	1.9 <sub>b</sub>	0.42	1.3 - 2.9
Verticaal	24	3.1	0.52	2.0 - 3.9	1.9 <sub>b</sub>	0.50	1.3 - 2.8

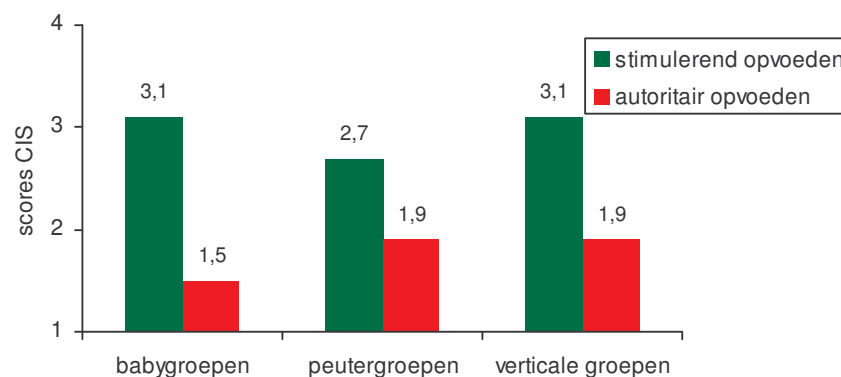
*Noot.* Gemiddelden met verschillende letters (verticaal) verschillen significant ( $p < .05$ ).

<sup>1</sup> Hier en in het vervolg van dit rapport zijn de dreumesgroepen ( $n = 2$ ) bij de babygroepen gevoegd.

Uit Tabel 4 kan worden afgeleid dat de leidsters in behoorlijke mate stimulerende opvoeding bieden (een gemiddelde van 3 op een schaal van 1 tot 4). Voor de subschaal ‘autoritair opvoeden’ wordt gemiddeld een 1.8 gescoord, wat betekent dat er in beperkte mate autoritair wordt opgevoed (schaalwaarde 1 = ‘helemaal niet’ en schaalwaarde 2 = ‘enigszins’). Om na te gaan in hoeverre de scores op de subschalen verschillen voor de drie typen groepen, is een multivariate variantie-analyse (MANOVA) uitgevoerd, met de gemiddelde scores op de subschalen ‘stimulerend’ en ‘autoritair’ als

afhankelijke variabelen, en type groep als onafhankelijke variabele. De resultaten lieten een hoofdeffect zien voor type groep (Pillais,  $F[4, 94] = 3.27$ ;  $p < .05$ ). Univariate tests wezen uit dat er alleen een significant verschil was voor de subschaal ‘autoritair opvoeden’ ( $F[2, 47] = 4.1$ ;  $p < .05$ ). Post hoc tests (Bonferroni) lieten zien dat de leidsters in de babygroepen significant lager scoorden op deze subschaal dan de leidsters in de peutergroepen en verticale groepen. Dit betekent dat in de babygroepen leidsters over het algemeen nog minder autoritair optreden dan in de overige groepen. In Figuur 10 worden de verschillende scores per type groep in een staafdiagram weergegeven.

Figuur 10: Scores op de subschalen van de CIS per type groep



### 3.3 Relaties tussen proceskwaliteit en structurele kenmerken

In deze paragraaf wordt nagegaan, in hoeverre verschillen in proceskwaliteit samenhangen met structurele kenmerken. Omdat de steekproef waarbinnen een ITERS-R is afgenomen te klein is ( $N = 15$ ) om als aparte groep te beschouwen binnen de correlatie- en regressie-analyses, die in het volgende gedeelte aan bod komen, zullen vanaf hier alle groepen uit de steekproef samengenomen worden. Hoewel de ITERS-R en ECERS-R niet identiek zijn, zijn de items en subschalen in voldoende mate gelijk, dat het gebruik van een gemiddelde score over de schalen heen gerechtvaardigd is (voor een inhoudsanalyse van de items en subschalen, zie paragraaf 4.1 en bijlage 3). Omdat gebleken is dat de scores op de ITERS-R over het algemeen wat hoger zijn dan de scores op de ECERS-R (zie Tabel 3), zullen we echter in de analyses de afname van het type schaal steeds als covariaat opnemen.

#### *Relaties tussen ITERS-R/ECERS-R en CIS*

Allereerst is nagegaan, in hoeverre de gemiddelde scores op de ITERS-R/ECERS-R samenhangen met de scores op de CIS. De Pearson correlaties zijn weergegeven in Tabel 5.

Tabel 5: Pearson correlaties tussen gemiddelde scores op de ITERS-R/ECERS-R en CIS

N = 50	CIS	
	Stimulerend	Autoritair
Totaal ITERS-R/ECERS-R	.53**	-.38**
<i>Subschalen</i>		
Ruimte/meubilering	.01	.07
Individuele zorg	.23	-.04
Taal	.67**	-.32*
Activiteiten	.06	-.34*
Interacties	.68**	-.43**
Programma	.40**	-.39**
Ouders en staf	.05	-.08

\*  $p < .05$ ; \*\*  $p < .01$

De subschalen ‘stimulerend opvoeden’ en ‘autoritair opvoeden’ van de CIS blijken naar verwachting samen te hangen met de totaalscore van de ITERS-R/ECERS-R. Een hogere totaalscore gaat samen met hogere scores van de leidster op de subschaal ‘stimulerend opvoeden’ ( $r = .53$ ), en met lagere scores op de subschaal ‘autoritair opvoeden’ ( $r = -.38$ ). Op subschaal niveau vallen vooral de correlaties tussen enerzijds ‘stimulerend opvoeden’ en anderzijds ‘interacties’ ( $r = .68$ ) en ‘taal’ ( $r = .67$ ) op. Dit betekent dat hogere scores op de subschalen waarbinnen interacties en taal een grote rol spelen, gepaard gaan met hogere scores van leidsters op ‘stimulerend opvoeden’ zoals gemeten met de CIS. In een volgende stap gingen we na in welke mate er verschillen zijn in ‘stimulerend opvoeden’ van leidsters in groepen die respectievelijk onvoldoende, matig en goed werden beoordeeld op de subschalen ‘taal’ en ‘interacties’ van de ITERS-R/ECERS-R (zie bijlagen 1 en 2). In Tabel 6 worden de gemiddelden en standaarddeviaties van de verschillende groepen weergegeven.

Tabel 6: Gemiddelde scores op de CIS gerelateerd aan de subschalen ‘taal’ en ‘interacties’ van de ITERS-R/ECERS-R

	ITERS-R/ECERS-R											
	Subschaal ‘taal’						Subschaal ‘interacties’					
	goed (n = 5)		matig (n = 30)		onvoldoende (n = 15)		Goed (n = 18)		matig (n = 23)		onvoldoende (n = 9)	
CIS	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
<i>Stimulerend</i>	3.73 <sub>a</sub>	.22	3.05 <sub>b</sub>	.45	2.57 <sub>c</sub>	.56	3.28 <sub>d</sub>	.52	3.01 <sub>d</sub>	.44	2.26 <sub>e</sub>	.29
95% CI	3.46 - 4.00		2.88 - 3.21		2.25 - 2.88		3.02 - 3.54		2.82 - 3.20		2.04 - 2.48	
<i>Autoritair</i>	1.83 <sub>f,g</sub>	.44	1.63 <sub>f</sub>	.29	2.07 <sub>g</sub>	.62	1.69 <sub>h</sub>	.36	1.67 <sub>h</sub>	.38	2.26 <sub>i</sub>	.56
95% CI	1.28 - 2.37		1.52 - 1.74		1.72 - 2.41		1.51 - 1.87		1.50 - 1.83		1.83 - 2.70	

*Noot.* Gemiddelden met verschillende letters (horizontaal) binnen het blok ‘taal’ en het blok ‘interacties’ verschillen significant ( $p < .01$ ); CI = confidence interval.

---

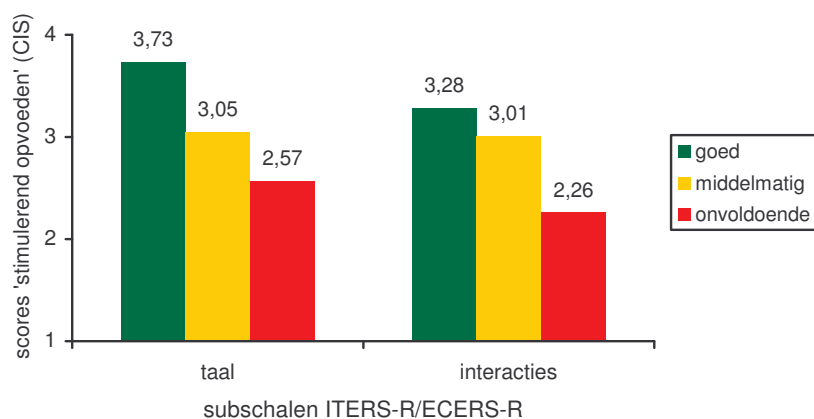
Een enkelvoudige variantie-analyse liet effecten zien voor de subschalen 'taal'  $F(2, 43) = 3.89, p = .03$  en 'interacties',  $F(2, 43) = 8.87, p = .00$  op 'stimulerend opvoeden'. Post hoc tests (Bonferroni) lieten zien dat voor de subschaal 'taal' de gemiddelde scores van de groepen die respectievelijk goed, middelmatig en onvoldoende scoorden, alle significant van elkaar verschilden. Voor 'interacties' verschilden de gemiddelde scores van de groepen die 'goed' versus 'onvoldoende' scoorden en van de groepen die 'middelmatig' versus 'onvoldoende' scoorden. Dit betekent dat leidsters uit groepen die op de ITERS-R/ECERS-R als goed (score  $\geq 5$ ) en middelmatig ( $3 \leq \text{score} < 5$ ) worden beoordeeld op de subschalen 'taal' en 'interacties', duidelijk meer stimulerend gedrag vertonen dan leidsters uit groepen die als onvoldoende (score  $< 3$ ) worden beoordeeld op de subschalen taal en interacties. Bovendien wordt dit verschil ook gevonden als we de vergelijking maken tussen 'goed' versus 'middelmatig' voor de subschaal 'taal'.

Ook op de subschaal 'autoritair opvoeden' bleken er effecten voor zowel 'taal',  $F(2, 43) = 3.50, p = .04$ , als 'interacties',  $F(2, 43) = 4.12, p = .02$ , te zijn. Post hoc tests (Bonferroni) toonden voor de subschaal 'taal' alleen een significant verschil aan tussen groepen die 'middelmatig' versus 'onvoldoende' scoorden. Leidsters uit groepen die 'onvoldoende' worden beoordeeld op 'taal', zijn geneigd meer autoritair op te treden dan de leidsters uit groepen die als 'middelmatig' worden beoordeeld. Voor de subschaal 'interacties' werd gevonden dat leidsters uit groepen die als 'onvoldoende' worden beoordeeld, een hogere score hebben op 'autoritair opvoeden' dan leidsters uit groepen die als 'goed' en 'middelmatig' worden beoordeeld.

In Tabel 6 zijn tevens voor alle gemiddelde waarden de 95% betrouwbaarheidsintervallen (CI = confidence interval) weergegeven. Hiermee wordt bedoeld dat de kans op een bepaald steekproefgemiddelde met 95% waarschijnlijkheid binnen de marge valt van de vermelde waarden. Uit deze betrouwbaarheidsintervallen valt af te leiden dat alle leidsters uit groepen die - hetzij op de subschaal 'taal' hetzij op de subschaal 'interacties' van de ITERS-R/ECERS-R - als 'goed' zijn beoordeeld, met 95% waarschijnlijkheid een score van 3 of hoger krijgen op de positieve subschaal ('stimulerend opvoeden') van de CIS. Leidsters daarentegen uit groepen die als 'onvoldoende' zijn beoordeeld op de betreffende subschalen van de ITERS-R/ECERS-R, krijgen met zeer grote waarschijnlijkheid een score die veel lager ligt dan 3 (tussen de 2.25 en 2.88 voor 'taal' en tussen de 2.04 en 2.48 voor 'interacties'). Hiermee wordt bevestigd dat een score van 3 of hoger op de subschaal 'stimulerend opvoeden' van de CIS als voldoende kan worden geclassificeerd.

In Figuur 11 worden de verschillen voor 'stimulerend opvoeden' in een staafdiagram weergegeven. Zoals hierboven al is uiteengezet, geldt voor de subschaal 'taal' dat de gemiddelden van alledrie de groepen significant van elkaar verschillen, terwijl voor de subschaal 'interacties' geldt dat de gemiddelde scores van leidsters uit groepen die als 'goed' en 'middelmatig' zijn geclassificeerd, significant hoger zijn dan de gemiddelde scores van leidsters die in de categorie 'onvoldoende' zijn geclassificeerd.

Figuur 11: Gemiddelde scores op 'stimulerend opvoeden' (CIS) voor drie categorieën groepen



#### *Relaties tussen ITERS-R/ECERS-R, CIS en structurele kenmerken*

Vervolgens is nagegaan in hoeverre er samenhang is tussen kenmerken van de groep, de leidsters en het kinderdagverblijf enerzijds (structurele kenmerken) en de gemiddelde scores op de ITERS-R/ECERS-R en de subschaal 'stimulerend opvoeden' van de CIS anderzijds. Om na te gaan wat de unieke invloed is van de verschillende typen structurele kenmerken op de ITERS-R/ECERS-R en positieve leidster-kind interacties ('stimulerend opvoeden'), zijn regressie-analyses uitgevoerd. Hierbij is uitgegaan van een theoretisch model (zie ook de Inleiding), waarin kenmerken die dichterbij het daadwerkelijk opvoeden binnen de groep liggen (proximale factoren) worden verondersteld een grotere invloed te hebben op de proceskwaliteit dan de zogenaamde distale factoren. Kenmerken die gerelateerd zijn aan de groep waarin het kind verblijft en aan de leidster zijn dus meer 'proximaal' dan kenmerken die gerelateerd zijn aan het kinderdagverblijf. Omdat het aantal structurele kenmerken te groot is om alle in een regressie-analyse op te nemen, hebben we op de volgende manier een selectie gemaakt. In ieder geval worden de drie variabelen groeps grootte, leidster-kind ratio en opleiding van de leidster - in de onderzoeksliteratuur vaak aangeduid als de 'ijzeren driehoek' (Tavecchio, 2002) - in de analyses opgenomen. Verder zal het type schaal worden opgenomen, omdat eerder in het rapport is aangetoond dat er over het algemeen hoger wordt gescoord door groepen bij welke een ITERS-R is afgenomen dan in groepen waarin een ECERS-R is afgenomen. Omdat eerder gerapporteerd is over een (geringe) samenhang tussen groepsindeling en proceskwaliteit (Van IJzendoorn et al., 1996; Gevers Deynoot-Schaub & Riksen-Walraven, 2002), zal ook de groepsindeling (horizontaal versus verticaal) in de analyses worden opgenomen. Voor de variabelen op leidsterniveau hebben we, naast het opleidingsniveau, een algemene variabele 'ervaring' opgenomen, op basis van de vier variabelen leeftijd, ervaringsjaren, ervaring op groep en uren aanstelling. De bivariate correlaties tussen deze variabelen varieerden van .34 tot .47; de interne consistentie (Cronbach's  $\alpha$ ) van de nieuwe factor was .59. Vervolgens zijn standaardscores berekend voor elke variabele afzonderlijk en is een nieuwe variabele 'ervaring' opgesteld, die de gemiddelde score weergeeft van de vier z-scores. Omdat in

eerder onderzoek (Gevers Deynoot-Schaub & Riksen-Walraven, 2002) naar voren is gekomen dat bestaansduur en het al of niet hebben van gesubsidieerde kindplaatsen gerelateerd is aan scores op de ITERS, hebben we in een additionele analyse deze twee variabelen in de laatste stap toegevoegd. In Tabel 7 worden de bivariate correlaties weergegeven van alle variabelen die in de regressie-analyses zijn opgenomen.

Uit Tabel 7 blijkt, dat er nauwelijks significante samenhangen zijn tussen structurele kenmerken enerzijds en proceskwaliteit anderzijds. Er zijn geen significante correlaties tussen de subschaal 'stimulerend opvoeden' van de CIS en de structurele kenmerken. Met betrekking tot de ITERS-R/ECERS-R scores wordt één significante correlatie weergegeven, namelijk met type schaal (ITERS-R of ECERS-R;  $r = -.37$ ). Deze correlatie is naar verwachting: Al eerder is gemeld dat er over het algemeen wat hoger gescoord wordt op de ITERS-R dan op de ECERS-R. Verder wordt er, naar verwachting, een significante correlatie gevonden tussen type schaal en groepsindeling ( $r = .54$ ): in de verticale groepen (op één na) is namelijk alleen een ECERS-R afgenomen.

Tabel 7: Pearson correlaties tussen structurele kenmerken en proceskwaliteit ( $N = 50$ )

	Proceskwaliteit		Groep				Leidster		Kdv	
	1	2	3	4	5	6	7	8	9	10
1. ITERS-R/ECERS-R	-									
2. CIS (stimulerend)	.53**	-								
3. Type schaal <sup>a</sup>	-.37**	-.14	-							
4. Groepsindeling <sup>a</sup>	-.10	.19	.54**	-						
5. Groepsgrootte	-.19	-.01	.28*	-.25	-					
6. Ratio	.20	-.08	-.30*	.14	-.65**	-				
7. Opleiding	.02	.04	.24	.26	-.02	-.04	-			
8. Ervaring	.19	.05	.07	.08	-.27	.09	.07	-		
9. Bestaansduur	.04	.14	-.11	-.11	-.09	-.04	.13	.34*	-	
10. Gesubsidieerd <sup>a</sup>	-.09	.07	.11	.04	.03	-.06	.20	-.10	-.18	-

*Noot.* <sup>a</sup> Deze variabelen zijn dichotoom gescoord (type schaal: 0 = ITERS-R, 1 = ECERS-R; groepsindeling 0 = horizontaal, 1 = verticaal; gesubsidieerd: 0 = ja, 1 = nee).

\*  $p < .05$ ; \*\*  $p < .01$

Er is een positieve samenhang tussen type schaal en groepsgrootte ( $r = .28$ ); dit wordt verklaard doordat in alle babygroepen een ITERS-R is afgenomen en babygroepen over het algemeen wat kleiner zijn dan peuter- en verticale groepen (zie paragraaf 3.1). Op dezelfde manier kan de negatieve samenhang ( $r = -.30$ ) tussen type schaal en leidster-kind ratio verklaard worden: Op babygroepen (ITERS-R) is de leidster-kind ratio over het algemeen hoger dan op peutergroepen en verticale groepen. De significante negatieve correlatie ( $r = -.65$ ) tussen groepsgrootte en leidster-kind ratio ligt voor de hand: Hoe groter de groep, des te lager de ratio over het algemeen is. De positieve samenhang

tussen ervaring van de leidster en bestaansduur van het kinderdagverblijf ( $r = .34$ ) kan verklaard worden doordat in de factor ‘ervaring’ de volgende drie componenten opgenomen zijn: leeftijd van de leidster, aantal jaren werkzaam in de kinderopvang, aantal jaren werkzaam op groep. Als leidsters over het algemeen ‘honkvast’ zijn en weinig van kinderdagverblijf wisselen, zal deze samenhang over het algemeen positief zijn. Om net als in de studie van 2001 na te kunnen gaan in hoeverre de ‘jongere’ kinderdagverblijven een lagere kwaliteit laten zien dan de ‘oudere’ kinderdagverblijven, is vervolgens de variabele ‘bestaansduur’ opgesplitst in ‘ $\leq 6$  jaar’ versus ‘ $> 6$  jaar’. Evenals in de studie van 2001 zijn er 18 kinderdagverblijven die 6 jaar of korter bestaan; de overige 32 kinderdagverblijven bestaan langer dan 6 jaar. Deze tweedeling liet geen verschillen in scores op de ITERS-R/ECERS-R zien ( $M = 3.13$  voor de ‘jongere’ kinderdagverblijven en  $M = 3.11$  voor de ‘oudere’ kinderdagverblijven).

De correlaties in Tabel 7 laten zien of en hoe de factoren 3 t/m 10 *afzonderlijk* samenhangen met de proceskwaliteit (ITERS-R/ECERS-R en CIS-stimulerend). Vervolgens zijn er meervoudige regressie-analyses uitgevoerd om te onderzoeken hoeveel de factoren 3 t/m 10 nu *samen* verklaren van de verschillen in proceskwaliteit tussen de kinderdagverblijven en wat de unieke bijdrage van elk van die factoren is. Als eerste is een hiërarchische lineaire regressie-analyse uitgevoerd met ITERS-R/ECERS-R score als afhankelijke variabele. In een eerste stap werden de variabelen op groepsniveau opgenomen en in een tweede stap de variabelen op leidsterniveau. De volgende 6 structurele kenmerken zijn in opeenvolgende stappen opgenomen in de regressie-analyse: (1) type schaal (ITERS-R of ECERS-R), groepsindeling (horizontaal versus verticaal), groepsgrootte, leidster-kind ratio, en (2) opleiding en ervaring leidster. De resultaten van deze analyse, nadat alle variabelen in de analyse zijn opgenomen, worden weergegeven in Tabel 8.

Tabel 8: Hiërarchische regressie-analyse op ITERS-R/ECERS-R scores

Variabelen	<i>B</i>	$\beta$	$sr^2$
Stap 1: Type schaal	-0.55	-.48*	
Groepsindeling	0.14	.13	
Groepsgrootte	0.02	.10	
Ratio	0.84	.09	.16
Stap 2: Opleiding	0.08	.08	
Ervaring	0.17	.22	.05
			$R^2 = .21$
			$R^2$ (aangepast) = .09
			$R = .45$

\*  $p < .05$ ; \*\*  $p < .01$

Uit Tabel 8 blijkt dat er geen significante model fit is na stap 2, met alle onafhankelijke variabelen in de vergelijking,  $R = .45$ ,  $F(6, 43) = 1.85$ ,  $p = .11$ . Na de eerste stap, met alleen de ‘groepsvariabelen’ in de regressie, is  $R^2 = .16$ ,  $F_{\text{toename}}(4, 45) = 2.06$ ,  $p = .10$ . Toevoeging van de leidstervariabelen in stap

2 leidt niet tot significante toename van de verklaarde variantie:  $R^2 = .21$ ,  $F_{\text{toename}}(6, 43) = 1.36$ ,  $p = .27$ . Zoals uit Tabel 8 blijkt, draagt alleen de variabele ‘type schaal’ significant bij aan de verklaarde variantie. Nadat gecontroleerd is voor het effect van ‘type schaal’, zijn er geen additionele effecten op groeps- en leidsterniveau. Dit betekent dus dat de kwaliteit van de opvang (ITERS-R/ECERS-R totaalscores) in dit onderzoek niet afhankelijk is van groepsindeling, groepsgrootte en leidster-kind ratio en evenmin van de opleiding en ervaring van de leidsters. In een laatste stap zijn de variabelen bestaansduur en gesubsidieerd opgenomen. Toevoeging van deze variabelen leidde echter niet tot een verbetering van de model fit,  $R = .46$ ,  $F(8, 41) = 1.39$ ,  $p = .23$ .

Eenzelfde regressie-analyse als hierboven beschreven is toegepast op de subschaal ‘stimulerend opvoeden’ van de CIS. De resultaten zijn weergegeven in Tabel 9. Er is geen significante model fit; na de laatste stap wordt slechts 5% van de variantie in CIS scores verklaard door de toegevoegde structurele kenmerken,  $R = .22$ ,  $F_{\text{toename}}(5, 44) = .45$ ,  $p = .82$ . Toevoeging van bestaansduur en subsidie leidt niet tot een betere fit,  $R = .29$ ,  $F_{\text{toename}}(7, 42) = .55$ ,  $p = .79$ . Geen van de structurele kenmerken draagt dus significant bij aan een verklaring van verschillen in scores op de CIS.

Tabel 9: Hiërarchische regressie-analyse op stimulerend opvoeden (CIS)

Variabelen	<i>B</i>	$\beta$	<i>sr</i> <sup>2</sup>
Stap 1: Groepsindeling	-0.22	-.20	
Groepsgrootte	0.00	-.03	
Ratio	-1.37	-.13	.05
Stap 2: Opleiding	-0.02	-.02	
Ervaring	0.03	.04	.00
			$R^2 = .05$
			$R^2$ (aangepast) = -.06
			$R = .22$

Vervolgens is onderzocht of er verschillen waren tussen groepen met een onvoldoende score ( $< 3$ ) en groepen met een middelmatige score ( $3 \leq \text{score} < 5$ ) op de ITERS-R/ECERS-R (zoals uit Figuur 9 blijkt, zijn er geen groepen die vallen in de categorie ‘goed’). Daartoe hebben we de invloed van dezelfde onafhankelijke variabelen onderzocht in een logistische regressie-analyse. De laatstgenoemde analysetechniek is gebruikt, omdat het onderscheid ‘middelmatig’ versus ‘onvoldoende’ in deze analyse de afhankelijke variabele (dichotoom) is. Zoals uit Tabel 10 blijkt, is er een significante model fit (discriminatie tussen groepen) op basis van alleen de vier groepsvariabelen (na stap 1),  $\chi^2(4, N = 50) = 12.01$ ,  $p = .02$ . Volgens de Wald test, bleek alleen ‘type schaal’ significant te differentiëren tussen beide groepen,  $z = 5.15$ ,  $p = .02$ . Na toevoeging van de leidstervariabelen (stap 2), was de model fit nog steeds significant, maar niet verbeterd,  $\chi^2(6, N = 50) = 15.23$ ,  $p = .02$ . Na toevoeging

van de structurele kenmerken bestaansduur en wel/niet gesubsidieerd, ontstond er geen significante verbetering van de model fit,  $\chi^2(6, N = 50) = 15.42, p = .05$ . De Wald test liet zien dat alleen de variabele ‘type schaal’ differentieert tussen groepen die laag respectievelijk gemiddeld scoren op de ITERS-R/ECERS-R. Zoals in Figuur 7 en 8 al is aangetoond, zijn er meer groepen waarbij een ECERS-R is afgenomen (54%) in de categorie ‘onvoldoende’ dan groepen waarbij een ITERS-R (7%) is afgenomen. Dit verklaart de hoge odds-ratio ( $Exp(B) = 27.14$ ) voor type schaal. De uitkomst van de logistische regressie-analyse betekent dat er vaker een onvoldoende score voor kwaliteit is toegekend in groepen waarin de kwaliteit is gemeten met de ECERS-R dan in groepen waarin de ITERS-R gebruikt is. Verder zijn er geen andere kenmerken die de groepen met een ‘middelmatige’ of ‘onvoldoende’ score voor kwaliteit van elkaar onderscheiden.

Tabel 10: Logistische regressie-analyse op gemiddelde versus lage kwaliteit (ITERS-R/ECERS-R)

Variabelen	<i>B</i>	<i>Exp(B)</i>	<i>Wald test</i>	<i>Model <math>\chi^2</math></i>
Stap 1: Type schaal	3.30	27.14	6.03*	
Groepsindeling	-.12	.90	.02	
Groepsgrootte	.11	1.12	.38	
Ratio	-1.67	.19	.05	12.01*
Stap 2: Opleiding	-.38	.69	.39	
Ervaring	.94	2.55	2.70	15.23*

\*  $p < .05$ ; \*\*  $p < .01$

### 3.4 Samenvatting kwaliteitsmeting 2005

Samengevat laat de landelijke peiling van 2005 zien dat de gemiddelde proceskwaliteit van de kinderdagverblijven - zoals gemeten met de ITERS-R/ECERS-R - op het minimum niveau zit, dat wil zeggen dat er opvang wordt geboden die amper voorziet in de stimulering van de basale ontwikkelingsmogelijkheden van kinderen. Geen enkele groep valt wat de totaalscore betreft in de categorie ‘goed’, 60% valt in de categorie ‘middelmatig’ en 40% in de categorie ‘onvoldoende’. In 20 groepen van de 50 kinderdagverblijven wordt opvang geboden die - volgens de normen van de ITERS-R/ECERS-R - niet voldoet aan de minimumeisen van de basale ontwikkelingsstimulering van kinderen. Op het niveau van subschalen worden de meeste onvoldoendes gegeven voor ‘zorg’ (90%) en ‘activiteiten’ (84%). De categorie ‘goed’ wordt het vaakst - als we de subschaal ‘staf’ buiten beschouwing laten - toegekend bij de subschaal ‘interacties’ (36%).

Er is een verschil geconstateerd in de proceskwaliteit zoals gemeten met de ITERS-R (babygroepen) versus de ECERS-R (peutergroepen en verticale groepen). In groepen waar de ITERS-R is gebruikt, is over het algemeen wat hoger gescoord dan in groepen waar de ECERS-R is toegepast. Wat de laatstgenoemde schaal betreft, werden geen verschillen geconstateerd tussen de gemiddelde

---

scores van peutergroepen en verticale groepen. De meeste ITERS-R groepen vallen in de categorie 'middelmatic' (93%), terwijl meer dan de helft (54%) van de ECERS-R groepen in de categorie 'onvoldoende' vallen. Bij de ITERS-R vallen voor 'zorg' alle groepen in de categorie 'onvoldoende'; bij de ECERS-R is dit percentage 86%. Op de subschaal 'activiteiten' wordt vooral onvoldoende gescoord door de ECERS-R groepen (94%). Ook als we naar de ITERS-R en ECERS-R afzonderlijk kijken, wordt het predikaat 'goed' relatief het vaakst toegekend voor de subschaal 'interacties' (respectievelijk 47% en 31%).

De leidsters laten kwalitatief voldoende interacties met de kinderen zien. Het blijkt dat de leidsters, als we afdaan op de gemiddelde scores op de CIS, in behoorlijke mate stimulerende ofwel responsieve opvoeding bieden en over het algemeen weinig autoritair optreden. Dit vonden we in alle typen groepen; alleen in de babygroepen werd nog iets minder autoritair opgetreden dan in de peutergroepen en verticale groepen. Hogere scores op de subschaal 'stimulerend opvoeden' blijken naar verwachting samen te hangen met scores op de ITERS-R en ECERS-R. Op subschaal-niveau blijkt deze samenhang vooral te bestaan voor 'taal' en 'interacties'. Leidsters uit groepen die voor de subschalen 'taal' en 'interacties' in de categorie 'goed' worden ingedeeld, worden hoger beoordeeld op 'stimulerend opvoeden' dan leidsters uit groepen die voor deze subschalen in de categorie 'onvoldoende' zijn ingedeeld. Wat de kern van het pedagogisch proces in de kinderopvang betreft wijzen onze gegevens in dezelfde richting: leidsters bieden over het algemeen de kinderen pedagogisch verantwoorde interacties. Buiten deze directe kern van interacties is de kwaliteit minder goed, en in een aantal opzichten zelfs slecht, bijvoorbeeld waar het hygiëne en verzorging betreft.

We vonden geen samenhangen tussen proceskwaliteit enerzijds en structurele kenmerken anderzijds. Als gecorrigeerd wordt voor type schaal (ITERS-R of ECERS-R) blijkt geen van de kenmerken op het niveau van groep, leidster en/of kinderdagverblijf significant bij te dragen aan de verklaring van verschillen in scores.



---

## 4. VERGELIJKING MET KWALITEITSMETINGEN UIT 1995 EN 2001

### 4.1 Vergelijkbaarheid

Voordat de scores op de verschillende kwaliteitsinstrumenten zullen worden vergeleken over de drie studies heen, gaan we na in hoeverre de drie studies vergelijkbaar zijn wat betreft steekproef (omvang, representativiteit), meetinstrumenten en structurele kenmerken.

#### *Steekproeven*

De omvang van de drie steekproeven is vergelijkbaar, respectievelijk  $N = 43$  groepen (uit 30 kinderdagverblijven),  $N = 51$  groepen (uit 39 kinderdagverblijven) en  $N = 50$  groepen (uit 50 kinderdagverblijven). In de steekproef uit 1995 was 21% van de groepen verticaal georganiseerd en 79% horizontaal. In 2001 waren deze percentages 37% verticaal en 63% horizontaal (alleen babygroepen en dreumesgroepen) en in 2005 48% verticaal ten opzichte van 52% horizontaal. Het percentage verticale groepen in de steekproeven is dus opgelopen over de jaren heen.

Zowel in 1995 als in 2005 is gebruik gemaakt van een aselechte landelijke steekproef. In de studie van 2001 werden echter alleen kinderdagverblijven in het westen en midden van Nederland benaderd. Om een indicatie te krijgen van eventuele verschillen in kwaliteit tussen kinderdagverblijven die al dan niet in de Randstad gesitueerd zijn, hebben we voor de steekproef van 2005 de gemiddelde proceskwaliteit per regio vergeleken. We hebben de steekproef opgesplitst in kinderdagverblijven in Noord- en Zuidholland, Utrecht en Flevoland ( $n = 29$ ) en daarbuiten ( $n = 21$ ). Er bleek geen verschil te zijn in gemiddelde scores voor de ITERS-R/ECERS-R tussen kinderdagverblijven binnen ( $M = 3.1$ ;  $SD = 0.52$ ) en buiten de Randstad ( $M = 3.2$ ;  $SD = 0.56$ ).

#### *Meetinstrumenten*

In de studie uit 1995 is gebruik gemaakt van de oorspronkelijke ITERS en ECERS; deze scores zijn samengevoegd tot een gecombineerde score. In de studie uit 2001 is gebruik gemaakt van de oorspronkelijke ITERS. De vraag dringt zich op in hoeverre de scores op de ITERS, ITERS-R, ECERS en ECERS-R psychometrisch en inhoudelijk vergelijkbaar zijn. Immers, als we vergelijkingen willen maken tussen de drie kwaliteitspeilingen, is het belangrijk na te gaan in hoeverre de herziene schalen hetzelfde meten als de oorspronkelijke schalen. De onderzoeksliteratuur over dit onderwerp is echter beperkt en de conclusies zijn niet eenduidig. Volgens een onderzoek uit de Verenigde Staten (Sakai, Whitebook, Wishard, & Howes, 2003), waarin zowel de ECERS als de ECERS-R bij dezelfde groepen in kinderdagverblijven is afgenomen, komen de gemiddelde scores, spreiding en psychometrische kwaliteiten van de ECERS-R overeen met die van het oorspronkelijke meetinstrument. Tietze et al. (2001) echter rapporteren in hun onderzoek bij 159 groepen, waarin de ECERS en ECERS-R gelijktijdig zijn afgenomen door dezelfde observator, een verschil van meer dan

---

een halve schaalpunt (.58) ten gunste van de ECERS. Hoewel er in in Duitsland dus aanwijzingen dat er op de ECERS-R wat lager wordt gescoord dan op de ECERS, zijn er in de Verenigde Staten geen bewijzen gevonden voor verschillen tussen de scores van de ECERS en ECERS-R (zie ook Clifford & Rossbach, in voorbereiding). Voor de ITERS-R is helaas (nog) niet zo'n studie verricht.

We vergelijken niet alleen de oorspronkelijke schalen met de herziene schalen, maar ook de ITERS(R) en ECERS(R) scores onderling. Om na te gaan in hoeverre deze vergelijking gerechtvaardigd is - op inhoudelijke gronden - hebben we aan de hand van een inhoudsanalyse de items en de daaraan gekoppelde subschalen van de vier schalen onderling vergeleken. Dit is ook gedaan voor de studie uit 1995, waarin overkoepelende subschalen zijn gemaakt voor de items van de ITERS en ECERS samen (voor een overzicht, zie bijlage 3). Uit deze inhoudsanalyse komt naar voren dat de items en subschalen van de ECERS het minst overeen komen met die van de andere schalen. Als we de ECERS even buiten beschouwing laten en we maken een vergelijking op item-niveau, blijkt dat over alle schalen heen minstens twee items in dezelfde subschaal zijn opgenomen. Uit bijlage 3 blijkt dat er voor 'ruimte/meubilering' vier items overeenkomen, voor 'individuele zorg' zes items, voor 'taal' twee items, voor 'activiteiten' acht items, voor 'interactie' drie items, voor 'programma' twee items en voor 'staf' vier items. De ECERS is met name een vreemde eend in de bijt als we kijken naar de subschalen 'interacties' en 'programma'. In de oorspronkelijke ECERS was geen subschaal 'interacties' opgenomen; de items uit 'social development' komen in de herziene schalen in verschillende subschalen terug. Items met betrekking tot 'programma' kwamen in de ECERS in het geheel niet voor.

In deze rapportage zal alleen een vergelijking gemaakt worden tussen de ITERS, ITERS-R en ECERS-R en gecombineerde ITERS/ECERS-scores. Immers, in de studie uit 1995 is gebruik gemaakt van gecombineerde subschalen, samengesteld uit items van zowel de ECERS als de ITERS (zie ook bijlage 3). In de studie uit 2001 is slechts gebruik gemaakt van de ITERS. Uit de inhoudsanalyse blijkt dat er in die mate overeenstemming is tussen de items van de verschillende schalen, dat een vergelijking - ook op subschaal-niveau - gerechtvaardigd is. Eerder in deze rapportage is geconstateerd dat er in 2005 hoger gescoord is op de ITERS-R dan op de ECERS-R. Voor de vergelijking met 2001, waarin alleen gebruik is gemaakt van de ITERS, zal hiermee rekening moeten worden gehouden. Overigens zijn voor alle schalen ook criteria ontwikkeld voor het onderscheid tussen goede, middelmatige, en onvoldoende kinderopvang. Deze onderscheiding is robuust omdat ze is gebaseerd op de totaalscores waarvoor individuele items (of het ontbreken ervan) niet veel gewicht in de schaal leggen. Vergelijkingen over de jaren heen in termen van goede, middelmatige en onvoldoende kwaliteit zijn dus betrouwbaarder dan vergelijkingen in termen van individuele items of subschalen.

#### *Structurele kenmerken*

In Tabel 11 wordt een overzicht gegeven van de structurele kenmerken die in minstens twee van de drie studies gemeten zijn. Voor de drie peilingen zijn gegevens beschikbaar met betrekking tot

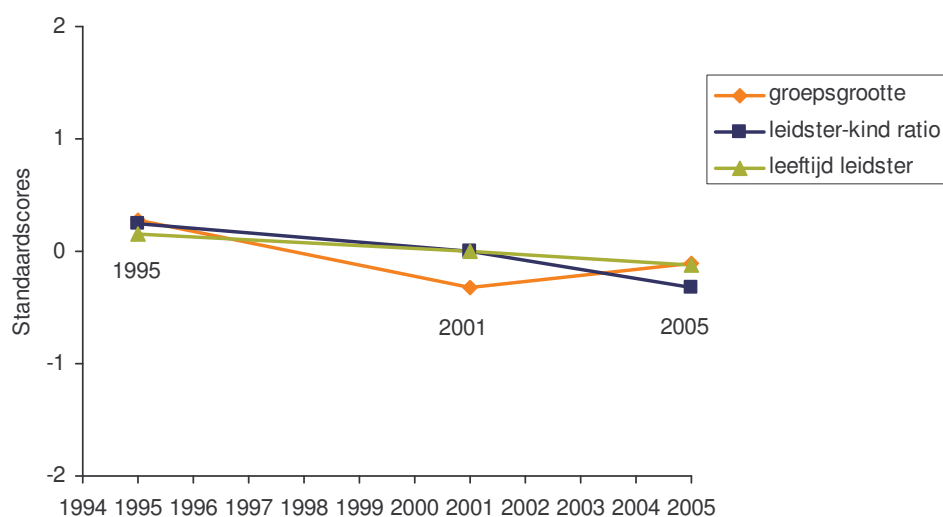
gemiddelde groeps grootte, leidster-kind ratio en leeftijd van de leidster. Zoals uit de tabel blijkt, ontlopen de drie studies elkaar niet veel wat betreft deze variabelen. De leidster-kind ratio blijkt echter in 2005 significant lager te zijn dan in 1995 ( $t(91) = 2.61; p < .01$ ) en in 2001 ( $t(99) = 2.84; p < .01$ ). Het aantal kinderen per leidster is in 2005 gemiddeld bijna 5 (4.9), terwijl deze aantallen in 1995 en 2001 gemiddeld 4.3 en 4.2 waren.

Tabel 11: Structurele kenmerken van de drie kwaliteitspeilingen

	1995		2001		2005	
	(N = 43)		(N = 51)		(N = 50)	
	M	SD	M	SD	M	SD
<i>Groep</i>						
Groeps grootte	11.1	4.0	9.1	2.6	9.7	2.6
Leidster-kind ratio	.27	.12	.26	.08	.22	.06
<i>Leidsters</i>						
Leeftijd	31.2	7.5	30.0	7.7	29.2	7.1
Ervaringsjaren in kinderopvang	5.5	3.5	-	-	6.4	4.5
Ervaringsjaren in groep	-	-	3.0	3.3	3.2	2.1
Uren aanstelling	30.0	6.4	-	-	30.0	6.1
<i>Kinderdagverblijven</i>						
Bestaansduur	-	-	9.9	5.6	10.0	6.8
			<i>N (%)</i>		<i>N (%)</i>	
Gesubsidieerd	-	-	27 (53%)		25 (50%)	

Om een zuivere grafische vergelijking te kunnen geven, hebben we de ruwe scores omgezet in standaardscores, met een gemiddelde van 0 en een gemiddelde standaarddeviatie van 1. Figuur 12 laat zien dat de drie studies vergelijkbaar zijn wat betreft leidster-kind ratio, groeps grootte en gemiddelde leeftijd van de leidster. Verder zijn voor twee studies vergelijkingen mogelijk op leidsterniveau - namelijk ervaringsjaren in de kinderopvang, ervaringsjaren in de groep en het aantal uren aanstelling per week - en op kinderdagverblijfniveau, namelijk bestaansduur en wel/niet gesubsidieerd. Ook deze variabelen ontlopen elkaar niet veel als we een vergelijking maken over de twee studies heen (zie Tabel 11). De laatste twee studies zijn vergelijkbaar als we kijken naar de gemiddelde bestaansduur (10 jaar) en het aantal centra dat gesubsidieerd werd: 27 (53%) in 2001 versus 25 (50%) in 2005.

Figuur 12: Vergelijking van structurele kenmerken tussen de drie peilingen



## 4.2 Proceskwaliteit

### *ITERS(R)/ECERS(R)*

Nagegaan is, in hoeverre er verschillen bestaan in proceskwaliteit als we de drie studies van het laatste decennium met elkaar vergelijken. In Tabel 12 wordt een overzicht gegeven van de gemiddelde scores en standaarddeviaties op de verschillende schalen.

Tabel 12: Vergelijking proceskwaliteit in 1995, 2001 en 2005

	1995		2001		2005		significante contrasten
	(N = 43)		(N = 51)		(N = 50)		
ITERS(R)/ECERS(R)	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Totaal	4.8	0.61	4.3	0.74	3.1	0.53	2005 < 2001 < 1995
<i>Subschalen</i>							
Ruimte/meubilering	5.1	0.98	4.9	1.07	3.4	0.66	2005 < 2001 en 1995
Individuele zorg	4.3	1.06	4.4	1.11	2.4	0.78	2005 < 2001 en 1995
Taal	4.7	1.01	3.7	1.30	3.4	1.04	2005 en 2001 < 1995
Activiteiten	4.3	0.82	3.6	0.71	2.4	0.66	2005 < 2001 < 1995
Interacties	5.3	0.97	5.1	1.46	4.3	1.35	2005 < 1995
Programma	5.2	1.24	-	-	3.9	0.98	2005 < 1995

*Noot.* In de laatste kolom wordt aangeven tussen welke jaren significante verschillen gevonden zijn;  $p < .003$  na Bonferroni correctie.

Uit Tabel 12 kan geconcludeerd worden dat de proceskwaliteit in 2005 gedaald is, zowel ten opzichte van 10 jaar geleden (1995) als ten opzichte van 4 jaar geleden (2001). Het verschil in gemiddelde

totaalscore tussen de jaren 1995 en 2001 komt overeen met een effectgrootte van  $d = 0.76$ , tussen 2001 en 2005 met een effectgrootte van  $d = 1.83$  en tussen 1995 en 2005 met een effectgrootte van  $d = 3.01$ . Dit zijn zeer sterke effecten. De effectgrootte Cohen's  $d$  is het gestandaardiseerde verschil tussen twee metingen. Cohen (1988) noemt  $d = 0.20$  een zwak effect,  $d = 0.50$  een middelmatig effect, en  $d = 0.80$  een sterk effect. Na selectie van de groepen uit 1995 waarbinnen een ITERS is afgenomen, worden met betrekking tot de vergelijking tussen 1995 en 2001 dezelfde verschillen in gemiddelde scores gevonden (Gevers Deynoot-Schaub & Riksen-Walraven, in druk).

Als we op subschaal-niveau een vergelijking maken van 2005 met 2001, is er een significante daling in kwaliteit te constateren voor de subschalen 'ruimte/meubilering', 'individuele zorg' en 'activiteiten'. Er is geen significante daling wanneer we kijken naar de subschalen 'taal' en 'interacties'. In de vergelijking met 10 jaar geleden is de kwaliteitsdaling merkbaar op alle subschalen. De bijbehorende  $t$ -waarden worden weergegeven in Tabel 13. Uit deze tabel valt tevens af te leiden tussen welke twee peilingen de verschillen statistisch significant waren.

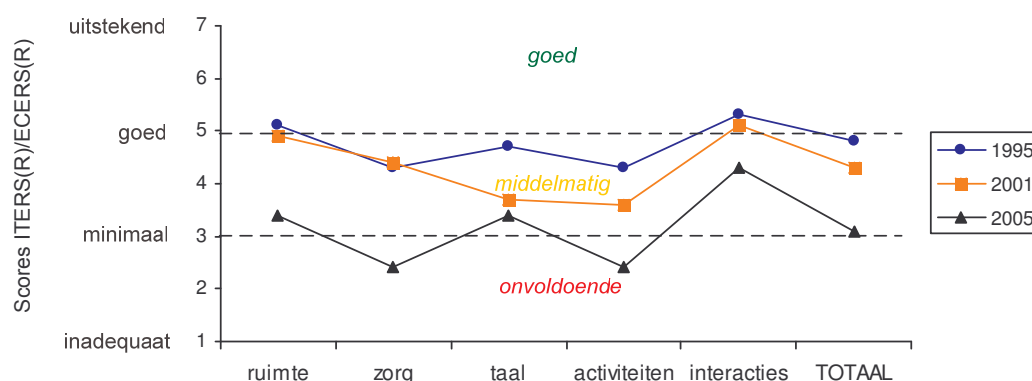
Tabel 13:  $t$ -waarden voor de verschillende toetsen

	1995 versus 2001 ( $df = 92$ )	2001 versus 2005 ( $df = 99$ )	1995 versus 2005 ( $df = 91$ )
	$t$	$t$	$t$
Totaal ITERS(R)/ECERS(R)	3.64*	9.08*	14.38*
<i>Subschalen</i>			
Ruimte/meubilering	1.18	8.26*	9.92*
Individuele zorg	0.49	10.16*	9.93*
Taal	4.00*	1.24	6.09*
Activiteiten	4.31*	8.44*	12.38*
Interacties	0.80	2.74	4.04*
Programma	-	-	5.64*

\*  $p < .003$  na Bonferroni correctie.

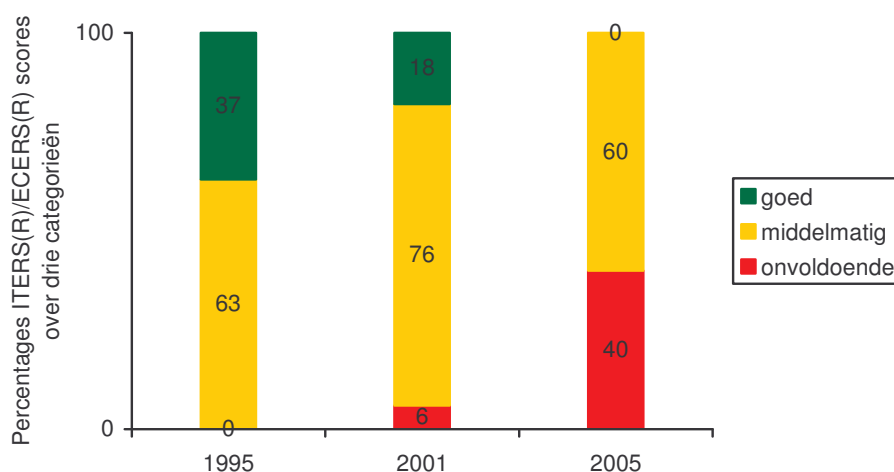
In Figuur 13 wordt de daling in proceskwaliteit (ITERS(R)/ECERS(R)) grafisch weergegeven aan de hand van kwaliteitsprofielen. Zoals uit deze figuur blijkt, is in 2005 is de gemiddelde kwaliteit van de onderdelen 'individuele zorg' en 'activiteiten' onvoldoende, van 'ruimte/meubilering' en 'taal' net voldoende en van het onderdeel 'interacties' ruim voldoende. In de jaren daarvoor was de gemiddelde kwaliteit op geen van de onderdelen onvoldoende.

Figuur 13: Trends in kwaliteitsprofielen over de jaren 1995, 2001 en 2005



Uit de gemiddelde scores is niet direct af te leiden in hoeveel gevallen er over de drie tijdstippen heen sprake is van opvang van goede, middelmatige en onvoldoende kwaliteit. We hebben onderzocht hoe de kinderdagverblijven in de drie steekproeven verdeeld zijn over de drie categorieën ‘goed’ (score  $\geq 5$ ), ‘middelmatig’ ( $3 \leq \text{score} < 5$ ) en ‘onvoldoende’ (score  $< 3$ ) (zie Figuur 14).

Figuur 14: Verdeling van de scores (in percentages) voor proceskwaliteit over de categorieën ‘goed’, ‘middelmatig’ en ‘onvoldoende’ over de jaren 1995, 2001 en 2005.

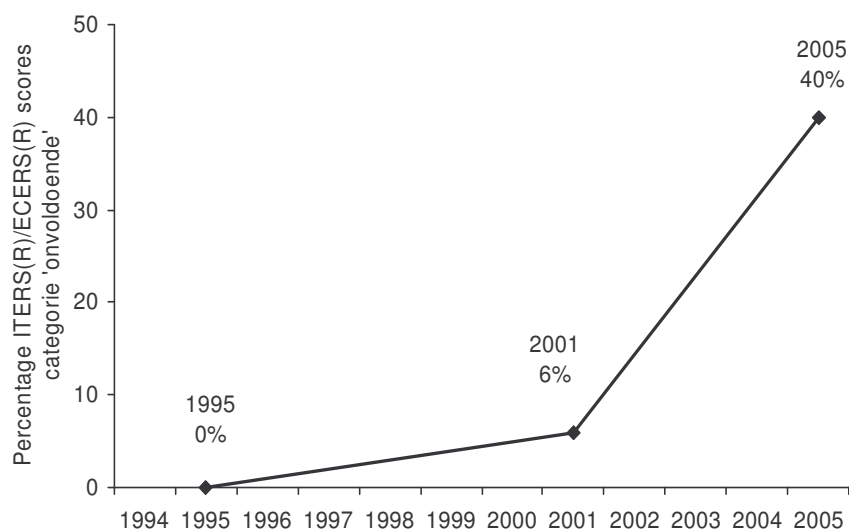


In 1995 werd in het geheel geen opvang van onvoldoende kwaliteit aangetroffen, in 2001 bood 6% van de centra onvoldoende kwaliteit en in 2005 is dit percentage opgelopen tot 40%. In 2001 bleken de meeste kwalificaties ‘onvoldoende’ voor te komen bij de subschaal ‘taal’ (31%) en ‘activiteiten’ (18%). Wat betreft ‘activiteiten’ wordt dit patroon herhaald in de studie van 2005; 60% van de groepen waarbij een ITERS-R is afgenomen scoort hierop onvoldoende en 94% van de groepen waarbij een ECERS-R is afgenomen (zie ook Figuren 7 en 8). In 2001 kreeg 18% van de kinderdagverblijven het predikaat ‘goed’, terwijl dit in 2005 geen enkele maal werd gegeven. Het

---

merendeel van de groepen wordt als 'middelmatig' geïnclassificeerd, zowel in 1995 (63%) als in 2001 (76%) en 2005 (60%). Figuur 15 laat de groei zien van het percentage groepen dat in de drie peilingen als 'onvoldoende' is beoordeeld.

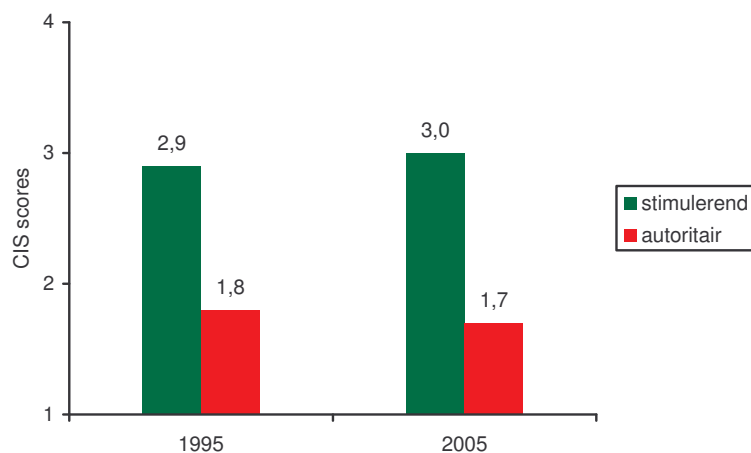
Figuur 15: Groei van het aantal groepen dat als 'onvoldoende' is geïnclassificeerd



### CIS

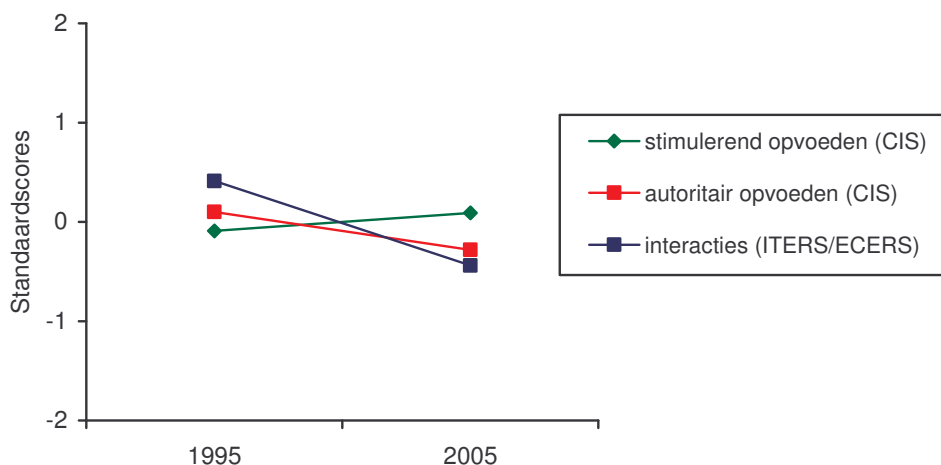
In 1995 is, net als in 2005, bij de leidsters een CIS afgenomen. De scores op de twee subschalen zijn vergelijkbaar voor beide jaren. De gemiddelde score op de subschaal 'stimulerend opvoeden' was 2.9 ( $SD = .55$ ) in 1995 tegenover 3.0 ( $SD = .57$ ) in 2005. Voor de subschaal 'autoritair opvoeden' lag het gemiddelde in 1995 op 1.8 ( $SD = .48$ ) en in 2005 op 1.7 ( $SD = .39$ ). Geconcludeerd kan worden dat de kwaliteit van de leidster-kind interacties, zoals gemeten met de CIS, stabiel is gebleven als we de studies uit 1995 en 2005 met elkaar vergelijken (zie Figuur 16). Helaas is in de studie uit 2001 geen CIS afgenomen, zodat we geen vergelijking kunnen maken over de drie studies heen.

Figuur 16: Scores op de CIS in 1995 en 2005



In Figuur 17 worden zowel de scores op de subschaal 'interacties' van de ITERS(R)/ECERS(R) als de scores op de subschalen van de CIS in één grafiek weergegeven. Om de metingen vergelijkbaar te maken, is gebruik gemaakt van standaardscores, met een gemiddelde van 0 en een standaarddeviatie van 1. Omdat er van de peiling van 2001 geen gegevens zijn van de CIS, is deze weggelaten uit de figuur. Zoals uit de figuur blijkt, is de kwaliteit van de leidster-kind interacties, zoals gemeten met de CIS, de afgelopen tien jaar zo goed als onaangetast gebleven. Wat betreft de interacties zoals gemeten met de subschaal van de ITERS(R) en ECERS(R) is een lichte daling te zien. In paragraaf 4.2 was al geconstateerd dat deze daling significant is, maar minder groot dan voor de andere subschalen.

Figuur 17: Kwaliteit van de leidster-kind interacties



---

### 4.3 Relaties tussen proceskwaliteit en structurele kenmerken

#### *Relaties tussen ITERS(R)/ECERS(R) en CIS*

Wat betreft de samenhang tussen de gemiddelde scores op de ITERS(R)/ECERS(R) en de CIS kan een vergelijking gemaakt worden tussen de studies van 1995 en 2005. De correlatie tussen de subschaal 'stimulerend opvoeden' van de CIS en de gemiddelde ITERS(R)/ECERS(R) score was .49 in de studie van 1995 en .53 in de studie uit 2005. De correlaties tussen 'autoritair opvoeden' en de gemiddelde ITERS(R)/ECERS(R) score waren voor beide studies gelijk, namelijk -.38. Voor beide peilingen geldt dus dat hogere scores op de ITERS(R) en ECERS(R) samengaan met hogere scores van de leidsters op 'stimulerend opvoeden' en lagere scores op 'autoritair opvoeden'. Opvallend is, dat in beide studies deze samenhangen het hoogst zijn voor de subschalen 'taal' (in 1995  $r = .42$  en  $r = -.40$ ; in 2005  $r = .67$  en  $r = -.32$ ) en 'interacties' (in 1995  $r = .61$  en  $r = -.47$ ; in 2005  $r = .68$  en  $r = -.43$ ). Geconcludeerd kan worden dat de scores op beide schalen voor proceskwaliteit een overeenkomstige samenhang vertonen tijdens beide peilingen. Overigens duiden de overeenkomstige correlaties met de CIS op de vergelijkbaarheid van de scores op de oorspronkelijke en de gereviseerde ITERS en ECERS, en is dit een bevestiging van onze inhoudsanalyses (zie paragraaf 4.1).

#### *Relaties tussen ITERS(R)/ECERS(R), CIS en structurele kenmerken*

Net als in 1995, hebben we in 2005 geen samenhangen gevonden tussen de scores op de CIS en structurele kenmerken. De relaties tussen structurele kenmerken en scores op de ITERS(R)/ECERS(R), zoals gerapporteerd in de studies van 1995 (leeftijd leidster, ervaring in de kinderopvang, aantal uren werkzaam en groepsindeling) en 2001 (groeps grootte, ervaring op de groep, groepsindeling, gesubsidieerd versus particulier, bestaansduur) zijn niet teruggevonden in de peiling van 2005. De relaties tussen kenmerken van het kinderdagverblijf, de leidsters en de groep enerzijds en proceskwaliteit anderzijds laten geen consistent beeld zien over de jaren heen.

### 4.4 Samenvatting vergelijking met voorgaande peilingen

Samengevat blijkt uit de vergelijking over de drie tijdstippen heen dat er sprake is van een significante daling in ITERS(R)/ECERS(R) scores. Deze daling is het meest zichtbaar voor de subschalen 'individuele zorg' en 'activiteiten'. In de vergelijking met 10 jaar geleden is de kwaliteitsdaling merkbaar op alle subschalen. Waar in 1995 in het geheel geen opvang van onvoldoende kwaliteit werd aangetroffen, is dit percentage in 2001 opgelopen tot 6% en in 2005 zelfs tot 40%. In 1995 kreeg 37% van de kindercentra het predikaat 'goed', in 2001 was dit percentage gedaald tot 18% en in 2005 zijn er geen kinderdagverblijven die in deze categorie vallen.

Deze daling in kwaliteit lijkt niet aan de orde voor de pedagogische kern van het werk in de kinderopvang, de interacties tussen leidsters en kinderen. De kwaliteit van deze interacties, zoals

---

gemeten met de CIS, is voldoende en stabiel gebleven ten opzichte van 10 jaar eerder. Bovendien is de teruggang in proceskwaliteit, zoals gemeten met de ITERS-R en ECERS-R, het minst uitgesproken voor de subschaal 'interacties'. In tegenstelling tot de jaren daarvoor, zijn er in 2005 geen significante samenhangen gevonden tussen structurele kenmerken enerzijds en proceskwaliteit anderzijds. Hoewel er in de steekproef van 2005 gemiddeld iets meer kinderen per leidster in de groepen aanwezig zijn, blijkt de leidster-kind ratio niet significant samen te hangen met proceskwaliteit.

---

## 5. CONCLUSIES EN DISCUSSIE

### *Belangrijkste conclusies*

De landelijke kwaliteitspeiling uit 2005 laat zien dat de gemiddelde proceskwaliteit van kinderdagverblijven - zoals gemeten met de ITERS-R/ECERS-R - weliswaar voldoet aan de meest basale criteria, maar tekort schiet als we de individuele ontwikkelingsmogelijkheden van kinderen in ogenschouw nemen. Op een schaal van 1 tot 7 wordt door Nederlandse kinderdagverblijven gemiddeld iets hoger dan een 3 gescoord. Geen enkele groep valt wat de totaalscore betreft in de categorie goed, 60% van de geobserveerde groepen valt in de categorie middelmatig en 40% in de categorie onvoldoende. In 2 op de 5 kinderdagverblijven wordt opvang geboden die niet voldoet aan de internationale maatstaven zoals in de ITERS-R en ECERS-R is vastgelegd. Geconstateerd is dat er verschillend gescoord is in groepen die beoordeeld zijn met de twee typen schalen. Op de ITERS-R (voor kinderen tot 2,5 jaar) worden over het algemeen iets hogere beoordelingen gegeven dan op de ECERS-R (voor kinderen van 2,5 tot 5 jaar). Vanwege het gebruik van verschillende meetinstrumenten in babygroepen versus peutergroepen en verticale groepen, kunnen daaraan geen uitspraken ontleend worden ten aanzien van eventuele verschillen in kwaliteit. Tussen peutergroepen en verticale groepen, in welke hetzelfde instrument was afgenomen, werden geen verschillen in proceskwaliteit geconstateerd.

Als we specifiek kijken naar subschalen en een relatieve vergelijking maken, blijkt dat de onderdelen ‘individuele zorg’ en ‘activiteiten’ het meeste reden baren tot zorg en het onderdeel ‘interacties’ het minst. Dit laatste wordt bevestigd door de resultaten van een observatie-instrument dat nader inzoomt op de leidster-kind interacties (de CIS): de leidsters blijken in behoorlijke mate stimulerende opvoeding te bieden en over het algemeen weinig autoritair op te treden. De kwaliteit van de interacties tussen leidsters en kinderen, zoals gemeten met de CIS, kan de vergelijking met 10 jaar geleden doorstaan; deze is voldoende en stabiel gebleven.

Voor een vergelijking over tijdstippen heen met betrekking tot de ITERS(R)/ECERS(R) geldt dit helaas niet. Er is sprake van een significante daling, zowel ten opzichte van de studie uit 1995 als ten opzichte van de studie uit 2001. De daling die al in 2001 was geconstateerd ten opzichte van 1995 heeft zich voortgezet. Op het niveau van subschalen is deze neerwaartse gang relatief het duidelijkst merkbaar op het gebied van ‘individuele zorg’ en ‘activiteiten’ en relatief het minst duidelijk voor ‘taal’ en ‘interacties’. De vergelijking over tijdstippen heen toont dus aan dat de kwaliteit van de leidster-kind interacties minder reden baart tot zorg als we de scores op de ITERS(R) en ECERS(R) in ogenschouw nemen en zeker niet zorgwekkend is als we de scores op de CIS vergelijken.

### *Mogelijke oorzaken*

De vraag dringt zich op, waarom Nederlandse kinderdagverblijven zo laag beoordeeld worden op de gehanteerde schalen voor proceskwaliteit, met name op aspecten van individuele zorg en activiteiten.

---

De schalen zijn ontwikkeld in de Verenigde Staten, waar relatief meer aandacht is voor hygiëne en de materiële omgeving dan in Nederland. In de schalen spelen kenmerken van de fysieke omgeving (individuele zorg, hygiëne, meubilering, ruimte en veiligheid) dan ook een relatief grote rol. Verder worden sommige items van de schalen gekenmerkt door een hiërarchie in indicatoren, waarin eerst voldaan moet zijn aan basale aspecten van hygiëne en de fysieke omgeving, alvorens andere indicatoren in de beoordeling meegenomen kunnen worden. Naar onze mening kunnen in sommige gevallen kenmerken van de fysieke omgeving gezien worden als noodzakelijke voorwaarden voor kwaliteit. Een groep die bijvoorbeeld regelmatig wordt verlaten door de leidsters zal een lage score krijgen op bepaalde items. Ongetwijfeld zullen leidsters uit deze groep - in ieder geval in kwantitatief opzicht - minder interacties hebben met kinderen. Anderzijds zijn er items verweven in deze schalen waarvan men zich kan afvragen in hoeverre het billijk is dat er binnen één item zowel indicatoren op het gebied van hygiëne als sociale interactie verweven zitten. Er zijn items waarbij de beoordeling blijft steken bij een 1 (inadequaat) omdat aan basale hygiënische procedures niet wordt voldaan. Het gevolg hiervan is dat indicatoren uit hogere schaalwaarden, waaraan wellicht wel wordt voldaan, niet meetellen in de beoordeling. Als de kinderen bijvoorbeeld hun handen niet hebben gewassen voor het eten, zullen de prettige sociale interacties tijdens het eten niet meetellen in de beoordeling. De lagere scores op 'activiteiten' kunnen verklaard worden doordat er over het algemeen te weinig ontwikkelingsmateriaal in de kinderdagverblijven aanwezig is (bijvoorbeeld blokken, materiaal voor fijne motoriek, materiaal voor rollenspel). Bovendien worden in de betreffende schalen voor een hogere beoordeling eisen gesteld aan de verscheidenheid in materialen en de toegankelijkheid. Als deze materialen er al zijn, is er vaak sprake van weinig variatie (bijvoorbeeld dezelfde type puzzels met dezelfde moeilijkheidsgraad) en/of mogen kinderen er niet vrij mee spelen. Verder wordt er in Nederland over het algemeen in kindercentra minder aandacht besteed aan 'voorschoolse stimulering' dan in de Verenigde Staten. Materialen op het gebied van bijvoorbeeld voorbereidend rekenen zijn in Nederland in de kleutergroepen (groepen 1 en 2) ruim vertegenwoordigd, maar niet in kinderdagverblijven. De vraag is overigens of een zwaarder accent op dergelijke vormen van vroegkinderlijke cognitieve ontwikkelingsstimulering ook wenselijk is.

Bovenstaande verklaart echter niet waarom kinderdagverblijven in de loop der jaren zoveel lager zijn gaan scoren op de gebruikte schalen, want onze maatstaven voor zorg en hygiëne zijn in die jaren niet drastisch veranderd. Omdat de opzet van dit onderzoek geen harde conclusies toelaat over mogelijke oorzaken van deze daling kunnen we hierover slechts speculeren. Gevers Deynoot-Schaub en Riksen-Walraven (in druk) noemen als mogelijke verklaringen voor de daling in kwaliteit tussen de jaren 1995 en 2001 de expansieve groei van kindercentra in die periode, met daaraan gekoppeld een tekort aan gekwalificeerd personeel en een hogere werkdruk. Als we de groei van de laatste vier jaar (zie Figuur 1) daarbij in ogenschouw nemen, is het aannemelijk dat deze verklaringen nog steeds stand houden. Een hogere werkdruk kan ten koste gaan van (een deel van) de processen die beoordeeld worden met het gehanteerde observatie-instrument. Van de leidsters wordt immers verwacht dat zij

---

zowel zorg dragen voor de gezondheid en fysieke veiligheid van de kinderen, als ook hun fysieke, taal-, intellectuele, en sociaal-emotionele ontwikkeling stimuleren. Ook de kwaliteit van de SPW-3 opleiding, die vrij algemeen is en niet speciaal toegesneden is op het werken met groepen jonge kinderen, kan als mogelijke verklaring gezien worden van de kwaliteitsdaling. Opvallend is, dat de huidige HKZ-certificering niet zonder meer kwaliteit garandeert. De groepen uit de kinderdagverblijven die in het bezit zijn van zo'n certificaat krijgen een nagenoeg zelfde beoordeling op de ITERS-R/ECERS-R ( $n = 13$ ;  $M = 3.18$ ) als de groepen die aangeven deze niet te bezitten ( $n = 37$ ;  $M = 3.10$ ). Van de 13 groepen van de kinderdagverblijven die in het bezit zijn van een HKZ-certificaat zijn er 4 die een onvoldoende beoordeling krijgen op de ITERS-R/ECERS-R. Blijkbaar meet de HKZ-certificeerder andere kwaliteitsaspecten dan in deze studie is gedaan.

Onduidelijk is, in hoeverre de doelgroep die gebruik maakt van hele dagopvang de laatste jaren veranderd is van samenstelling. Een beperking van de beschreven kwaliteitspeiling is dat er geen achtergrondinformatie is verzameld over de kinderen en dat er wat betreft de populatie kinderen dus geen vergelijking gemaakt kan worden met voorgaande peilingen. Indien het aandeel kinderen met een lagere sociaal-economische status in de hele dagopvang gegroeid is, zou dit wellicht gedeeltelijk bij kunnen dragen aan een verklaring van de daling in kwaliteit. Deze kinderen vragen namelijk over het algemeen meer aandacht omdat zij minder 'bagage' (bijvoorbeeld een achterstand in taalontwikkeling, cognitieve ontwikkeling) van huis uit meekrijgen. Vanuit het onderwijs is immers ook bekend, dat scholen met een relatief grote populatie kinderen met een lagere sociaal-economische status over het algemeen gemiddeld kwalitatief onder de maat zijn. In de kinderopvang staat weliswaar de cognitieve ontwikkeling niet op de voorgrond, maar het is zeer goed voorstelbaar dat er een groter beroep wordt gedaan op de vaardigheden van leidsters als er meerdere kinderen in de groep zitten die meer dan gemiddelde aandacht behoeven. Voor de steekproeven van de landelijke peilingen zijn deze gegevens echter niet beschikbaar, zodat over een eventuele bijdrage van deze factor aan de daling in kwaliteit slechts gespeculeerd kan worden. Het is in elk geval van belang om in toekomstig onderzoek achtergrondinformatie (sociaal-economische status, etniciteit, en dergelijke) over de kinderen die gebruik maken van kinderopvang op te nemen.

### *Beperkingen*

Dit onderzoek kent een aantal methodologische beperkingen, waardoor voorzichtigheid is geboden bij het interpreteren van de resultaten. Ten eerste wijzen we erop dat er in de drie studies sprake is van verschillende steekproeftrekkingen; de groepen verschillen zowel wat leeftijd van de kinderen als wat geografische spreiding betreft. Ten tweede attenderen we op het gebruik van gelijksoortige, maar niet identieke schalen in de drie studies. Omdat er in de verschillende studies schalen zijn gebruikt die weliswaar zeer sterke verwantschap vertonen, maar niet identiek zijn, is voorzichtigheid geboden bij het interpreteren van de omvang van de daling. Uit onderzoeksliteratuur uit de Verenigde Staten, (Sakai et al.; zie paragraaf 4.1) blijkt dat de gemiddelde scores van de ECERS-R overeenkomen met

---

die van de ECERS. In een Duits onderzoek (Tietze et al. 2001) wordt echter melding gemaakt van meer dan een half schaalpunt verschil in het voordeel van de ECERS. Omdat de conclusies niet eenduidig zijn en gebaseerd zijn op een beperkt aantal studies, is vooralsnog onduidelijk of de geconstateerde verschillen in kwaliteit, zoals gemeten met de oorspronkelijke en met de herziene versie, voor 100% 'daadwerkelijke' verschillen representeren. Wellicht kan een (heel klein) deel op het conto geschreven worden van de herzieningen, waarbij een iets andere manier van scoren wordt gehanteerd en informatie met betrekking tot het scoren toegankelijker is geworden (onder andere via website en handboek). Een mogelijk effect hiervan is dat er 'strenger' wordt gescoord. In de hier beschreven kwaliteitspeiling is het verschil met voorgaande jaren echter zo groot (voor de totaalscores 1.2 schaalpunt ten opzichte van 2001 en 1.7 schaalpunt ten opzichte van 1995), dat dit niet louter door revisies en/of het gebruik van schalen voor jongere versus oudere kinderen verklaard kan worden en er dus zeker sprake is van een daling. Op het gebied van de ITERS en ITERS-R is er nog nooit een vergelijkende studie geweest. Wat betreft de vergelijking tussen de ITERS en ECERS, lieten Gevers Deynoot-Schaub en Riksen-Walraven (in druk) zien dat in de steekproeven uit 1995 en 2001 de gemiddelde scores op beide schalen vergelijkbaar zijn.

Afgezien van de discussie of we de daling als extreem, fors of tamelijk groot duiden, blijft het gegeven staan dat volgens internationale maatstaven Nederlandse kinderdagverblijven anno 2005 gemiddeld een laag niveau van kwaliteit hebben. De resultaten van recente internationale studies laten zien dat de gemiddelde scores van de groepen uit de peiling van 2005 achterblijven bij die van bijvoorbeeld de Verenigde Staten, Canada en Duitsland. In de Verenigde Staten worden op de ECERS-R gemiddelde scores gerapporteerd van rond de 3.90 (Fiene et al., 2002) en 4.87 (Sakai et al., 2003). In een groot Canadees onderzoek lag het gemiddelde op 4.70 (Goelman et al., 2000) en ook in Duitsland zijn gemiddelde scores vermeld van 4.20 (Erning, 2003) en 4.19 (Tietze et al., 2001). Dit is een aanzienlijke verslechtering van onze internationale positie in vergelijking tot de meting uit 1995 toen we nog konden concluderen dat "the quality of center care in the Netherlands was rather good compared with that of other European and North American countries" (Van IJzendoorn et al., 1998, p. 437). We moeten dan ook vaststellen dat de Nederlandse kinderopvang haar koppositie in de internationale vergelijkingen is kwijtgeraakt.

De kwaliteitsmeting uit 2005 mag niet gezien worden als een peiling die de stand van zaken in kaart brengt vóórdat de Wet kinderopvang in werking is getreden (een zogenaamde 'nulmeting'). Ten eerste was de nieuwe Wet kinderopvang al in werking getreden op het moment dat de dataverzameling een aanvang nam. Bovendien is de invoering van de wet een jaar uitgesteld; deze zou eigenlijk op 1 januari 2004 ingaan. De meeste kindercentra hebben daardoor ruim de tijd gehad zich op de eisen en mogelijkheden van de nieuwe wetgeving voor te bereiden en hun beleid en praktijk alvast aan te passen. Als een gebeurtenis niet plotseling uit het niets opdoemt, zoals bij de Wet kinderopvang, wordt de interpretatie van een mogelijk effect nog eens extra bemoeilijkt. Er mag namelijk niet onderschat worden wat een nieuwe wet in de aanloopfase teweeg kan brengen bij alle betrokkenen van de

---

kinderopvang. Ten slotte is het lastig via onderzoek vast stellen wat de bijdrage is van een externe, niet te beïnvloeden factor - in dit geval de Wet kinderopvang - op de kwaliteit van kinderopvang zonder een lange reeks op gelijke wijze opgezette kwaliteitsmetingen voorafgaand aan en na invoering van zo'n externe factor, zodat een zogenaamde trendanalyse mogelijk wordt. Voor een trendanalyse zijn meer meetpunten in de tijd nodig dan nu het geval is. We kunnen alleen constateren dat de dalende trend in de opvangkwaliteit tussen 1995 en 2001 zich heeft voortgezet in de jaren daarna.

#### *Aanbevelingen voor toekomstig onderzoek*

We pleiten ervoor kwaliteitspeilingen te verrichten bij grotere steekproeven. In tegenstelling tot in 1995 en 2001, zijn er in 2005 geen samenhangen gevonden tussen proceskwaliteit enerzijds en structurele kenmerken anderzijds. Als gecorrigeerd wordt voor type schaal (ITERS-R of ECERS-R) blijkt geen van de kenmerken op het niveau van groep, leidster en/of kinderdagverblijf significant bij te dragen aan de verklaring van verschillen in scores. De inconsistenties in het al dan niet vinden van significante samenhangen, kunnen (deels) verklaard worden door de kleine steekproeven. Voor het adequaat bestuderen van samenhangen tussen structurele kenmerken enerzijds en proceskwaliteit anderzijds, zijn grotere steekproeven nodig die voldoende differentiëren tussen structurele kenmerken op groeps-, leidster- en kinderdagverblijfniveau. Kleine, maar praktisch relevante, samenhangen kunnen in steekproeven van 50 groepen vanwege een tekortschietende statistische 'power' niet goed worden vastgesteld. Verder is het mogelijk dat de geringe spreiding in scores, zowel wat betreft de structurele kenmerken als de proceskwaliteit, een rol speelt bij het niet vinden van samenhangen.

Een laatste maar niet onbelangrijk punt dat we willen aanstippen is dat de kwaliteit van de interacties tussen leidsters en kinderen ondervertegenwoordigd is in deze peiling (en de peilingen hiervoor). In dit onderzoek en ook in de peiling van 10 jaar geleden is naar voren gekomen dat de leidster-kind interacties, zoals gemeten met de CIS, voldoende zijn. De CIS geeft echter slechts een globale en tegelijk beperkte indicatie van de kwaliteit van de leidster-kind interacties. Geconcludeerd kan worden dat leidsters over het algemeen voldoende sensitief en positief zijn in hun interacties met kinderen. Het is echter geen geschikt instrument voor het in kaart brengen van alle interactievaardigheden van leidsters, die bevorderlijk zijn voor de cognitieve ontwikkeling of taalontwikkeling van jonge kinderen. In ontwikkelingspsychologisch onderzoek is aangetoond dat een leidster die met een groep jonge kinderen werkt in ieder geval de volgende vaardigheden in haar interacties met jonge kinderen zou moeten laten zien: sensitieve responsiviteit, respect voor autonomie, structureren en grenzen stellen, praten en uitleggen en het begeleiden van de interacties tussen kinderen (Riksen-Walraven, 2004). De CIS meet vooral sensitieve responsiviteit, maar de andere leidstervaardigheden komen niet of nauwelijks aan bod. In toekomstig onderzoek zal specifieke aandacht besteed moeten worden aan deze vaardigheden van leidsters in de dagelijkse omgang met de kinderen. Ook de gedragingen van kinderen tijdens het verblijf op het kinderdagverblijf - het welbevinden en de betrokkenheid - zien wij als belangrijke indicatoren van pedagogische kwaliteit en

---

zullen daarom in een kwaliteitsinstrument opgenomen moeten worden. In een volgende studie zal het NCKO, met een compleet instrument dat alle relevante indicatoren in kaart brengt (zie ook de Inleiding), de basis kunnen leggen voor toekomstige kwaliteitsmetingen. Een kwaliteitsmeting zoals hierboven beschreven, waarin *alle* aspecten van pedagogische kwaliteit vertegenwoordigd zijn, legt niet alleen de basis voor toekomstige kwaliteitsmetingen, maar tevens voor gerichte interventies ter verbetering van de pedagogische kwaliteit van kinderdagverblijven.

#### *Aanbevelingen voor praktijk en beleid*

Als we willen dat kinderopvang meer is dan louter ‘opvang’ die voldoet aan de minimum kwaliteits-eisen, zal geïnvesteerd moeten worden in de pedagogische kwaliteit van kinderopvang in Nederland. De resultaten van dit onderzoek laten zien dat het merendeel van de geobserveerde groepen uit de steekproef onder de maat beoordeeld wordt op de aspecten ‘individuele zorg’ en ‘activiteiten’. Extra investeringen, zowel vanuit de praktijk als vanuit het beleid, zullen zich dan ook in eerste instantie op deze aspecten moeten richten. Deze investeringen mogen echter niet ten koste gaan van de kern van het pedagogisch proces binnen de kinderopvang, namelijk de kwaliteit van de leidster-kind interacties. Op basis van deze overweging stelt het NCKO de volgende aanbevelingen voor.

Ten eerste pleiten wij ervoor dat kinderdagverblijven er zorg voor dragen dat voldaan wordt aan de basale eisen van hygiëne en veiligheid. We hebben regelmatig geobserveerd dat leidsters en kinderen hun handen niet wassen met zeep, dat er geen schone slaapplekken zijn en/of dat kinderen deze moeten delen, dat er geen adequate supervisie is en dat de veiligheid van de binnen- en/of buitenruimtes te wensen overlaat. Een basisvoorwaarde voor kwaliteit - zeker geredeneerd vanuit het standpunt van ouders en vanuit het perspectief van de volksgezondheid - is dat deze aspecten op een aanvaardbaar minimumniveau zitten. De eisen, zoals gehanteerd door de GGD, vormen hierbij een belangrijk aanknopingspunt.

Ten tweede zou er meer aandacht moeten komen voor pedagogische activiteiten in een verantwoorde en stimulerende omgeving. Dit vergt zowel een materiële als een organisatorische investering. Er kan het nodige verbeterd worden aan de kindvriendelijkheid van de ruimte en meubilering. Hierop worden de groepen over het algemeen laag beoordeeld omdat de leidsters vanwege de inrichting geen overzicht hebben over alle kinderen, er vaak te weinig duidelijk herkenbare en handig uitgeruste speelhoeken zijn en er geen kindgerichte aankleding is (de wanden zijn vaak kaal, er hangt geen/weinig werk van kinderen, platen hangen op ooghoogte van volwassenen). Adequate, op de ontwikkeling van de kinderen afgestemde materialen moeten zowel talrijker als gevarieerder aanwezig zijn. Dit geldt met name voor blokken, muziekinstrumenten, materiaal voor rollenspel, boeken/materialen over de natuur en knutselmaterialen. Deze materialen moeten op ooghoogte staan van de kinderen en tijdens het spel vrij toegankelijk zijn. Om acceptatie van verscheidenheid tussen mensen te bevorderen zouden de materialen deze verscheidenheid moeten laten zien (bijvoorbeeld platen, puzzels, en boeken met afbeeldingen van mensen van verschillende

---

rassen, culturen, leeftijden). Voor de leidsters is een belangrijke rol weggelegd bij het begeleiden van activiteiten met deze materialen en het stimuleren van de ontwikkeling van de kinderen, bijvoorbeeld door middel van praten en uitleggen. Met de toenemende groei van verticale groepen is dit een investering die zeker de aandacht vraagt, omdat de omgeving en het materiaal afgestemd moeten worden op de ontwikkeling van zowel baby's als peuters.

Als laatste stellen wij voor groepsassistenten aan te stellen ter ondersteuning van de pedagogische leidsters. De werkdruk van leidsters is hoog en het is niet onze intentie deze met bovengenoemde aanbevelingen nog extra te verhogen. Van leidsters wordt verwacht dat zij zowel verzorgende als pedagogische taken combineren met een grote groep kinderen onder hun hoede. De groepsassistent kan deze taak verlichten door de basale verzorgende aspecten uit handen te nemen van de leidsters. Zij zet bijvoorbeeld de lunch klaar, zorgt dat de kinderen hun handen hebben gewassen en zet de materialen klaar voor een knutselactiviteit. De leidster kan zich daardoor volledig richten op haar pedagogische taak. Het inzetten van groepsassistenten verlicht niet alleen de werkdruk van de leidster, maar vergroot bovendien de status van het beroep alsmede het werkplezier.



---

## REFERENTIES

- Arnett, J.(1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.
- Clifford, R.M., & Roßbach, H. (in voorbereiding). *Structure and stability of the Early Childhood Environment Rating Scale*.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum.
- Cryer, D., Tietze, W., Burchinal, M., Leal, T., & Palacios, J. (1999). Predicting process quality from structural quality in preschool programs: A cross-classroom comparison. *Early Childhood Research Quarterly, 14*, 339-361.
- Cryer, D., Wagner-Moore, L., Burchinal, M., Yazejian, N., Hurwitz, S., & Wolery, M. (2005). Effects of transitions to new child care classes on infant/toddler distress and behavior. *Early Childhood Research Quarterly, 20*, 37-56.
- De Kruif, R.E.L., & Tavecchio, L.W.C. (2004). Instrumenten voor meting van kwaliteit van kinderopvang: beschrijving en evaluatie. In M.H. van IJzendoorn, L.W.C. Tavecchio & J.M.A. Riksen-Walraven (Red.). *De kwaliteit van de Nederlandse kinderopvang* (pp. 124-153). Amsterdam: Boom.
- De Schipper, J.C. (2003). *Children in flexible child care: Experiences of stability in center care, quality of care and the child's adjustment*. Dissertatie: Universiteit Leiden.
- Erning, G. (2003). *Qualitätsentwicklung in Kindergarten: Abschlußbericht der Fortbildung 2001-2003*. Bamberg: Otto-Friedrich Universität.
- Fiene, R., Greenberg, M., Bergsten, M., Fegley, C., Carl, B., & Gibbons, E. (2002). *Early care and education: The keystone of Pennsylvania's future*. Pennsylvania State University.
- Gevers Deynoot-Schaub, M., & Riksen-Walraven, M. (2002). Kwaliteit onder druk: De kwaliteit van Nederlandse kinderdagverblijven in 1995 en 2001. *Pedagogiek, 22*, 109-124.
- Gevers Deynoot-Schaub, M., & Riksen-Walraven, M. (in druk). Child care under pressure: The quality of Dutch centers in 1995 and 2001. *The Journal of Genetic Psychology*.
- Goelman, H., Doherty, G., Lero, D.S., LaGrange, A., & Tougas, J. (2000). *You bet I care! Caring and learning environments: Quality in child care centres across Canada*. Ontario: Centre for Families, Work and Well-being.
- Harms, T., Clifford, R.M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale- Revised*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Environment Rating Scale- Revised*. New York: Teachers College Press.
- Peisner-Feinberg, E.S., & Burchinal, M.R. (1997). Relations between preschool children's child-care experiences and concurrent development: The Cost, Quality, and Outcome Study. *Merrill-Palmer Quarterly, 43*, 451-477.

- 
- Riksen-Walraven, J.M.A. (2004). Pedagogische kwaliteit in de kinderopvang: doelstellingen en kwaliteitscriteria. In M.H. van IJzendoorn, L.W.C. Tavecchio & J.M.A. Riksen-Walraven (Red.), *De kwaliteit van de Nederlandse kinderopvang* (pp. 100-123). Amsterdam: Boom.
- Sakai, L.M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly, 18*, 427-445.
- Tavecchio, L.W.C. (2002). Kinderopvang: Opvoedingsmilieu tussen gezin en school. In M.H. van IJzendoorn & H. de Frankrijker (Red.), *Pedagogiek in beeld: Een inleiding in de pedagogische studie van opvoeding, onderwijs en hulpverlening* (pp. 137-153). Houten/Diegem: Bohn Stafleu Van Loghum.
- Tavecchio, L.W.C., Riksen-Walraven, J.M.A., & Van IJzendoorn, M.H. (2002). *Nationale Studie Pedagogische Kwaliteit Kinderopvang: Onderzoeksvoorstel*.
- Tietze, W., Bethke, C., Braukhane, K., Roßbach, H., Schlecht, D., & Schmidt, N. (2001). *Analyse und Sicherung pädagogischer Qualität in den Tageseinrichtungen der Bremischen Evangelischen Kirche*. Berlin: Freie Universität.
- Tietze, W., Cryer, D., Bairrão, J., Palacios, J., & Wetzel, G. (1996). Comparison of observed process quality in early child care and education programs in five countries. *Early Childhood Research Quarterly, 11*, 447-475.
- Van der Kemp, S., & Kloosterman, M. (2005). *Het aanbod van kinderopvang per eind 2004*. Leiden: Research voor Beleid.
- Van IJzendoorn, M.H., Tavecchio, L.W.C., Verhoeven, M.J.E., Reiling, E.J., & Stams, G.J.J.M. (1996). De kwaliteit van de Nederlandse kinderopvang. *Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs, 12*, 286-313.
- Van IJzendoorn, M.H., Tavecchio, L.W.C., Stams, G.J.J.M., Verhoeven, M.J.E., & Reiling, E.J. (1998). Quality of center day care and attunement between parents and caregivers: Center day care in cross-national perspective. *The Journal of Genetic Psychology, 159*, 437-454.
- Vervoort, Y., Vinders, S., & MacLennan, W. (2003). *Monitor uitbreiding kinderopvang 2003*. Netwerkbureau Uitbreiding Kinderopvang.

---

## BIJLAGE 1

### Overzicht van de subschalen en items van de ITERS-R

39 items in 7 subschalen

#### **Ruimte en Meubilering**

1. Binnenruimte
2. Meubilering voor dagelijkse verzorging en spel
3. Voorzieningen voor ontspanning en comfort
4. Indeling van de ruimte
5. Aankleding van de ruimte voor kinderen

#### **Dagelijkse individuele verzorging**

6. Brengen en halen
7. Maaltijden en tussendoortjes
8. Slapen
9. Verschonen/toiletgebruik
10. Gezondheidspraktijk
11. Veiligheidspraktijk

#### **Luisteren en praten**

12. Kinderen helpen taal te begrijpen
13. Kinderen helpen taal te gebruiken
14. Het gebruik van boeken

#### **Activiteiten**

15. Fijne motoriek
16. Actief lichamelijk spel
17. Knutselen
18. Muziek en beweging
19. Blokken
20. Rollenspel
21. Zand- en waterspel
22. Natuur/wetenschap
23. Gebruik van tv, video, en/of computers
24. Bevorderen van het accepteren van verscheidenheid

#### **Interacties**

25. Toezicht op ontwikkeling en spel
26. Interacties tussen kinderen
27. Leidster-kind interacties
28. Discipline (handhaven van orde)

#### **Programma**

29. Dagschema/dagindeling
30. Vrij spel
31. Spelactiviteiten in een groep
32. Voorzieningen voor kinderen met een handicap

#### **Ouders en Staf**

33. Voorzieningen voor ouders
34. Voorzieningen voor de persoonlijke behoeften van de staf
35. Voorzieningen voor de professionele behoeften van de staf
36. Samenwerking en interactie tussen leidsters
37. De contuïteit van de staf
38. Toezicht op en evaluatie van de staf
39. Mogelijkheden voor professionele ontwikkeling



---

## BIJLAGE 2

### Overzicht van de subschalen en items van de ECERS-R

43 items in 7 subschalen

#### Ruimte en Meubilering

1. Binnenruimte
2. Meubilering voor verzorging, spel en ontwikkeling
3. Meubilering voor ontspanning en gerief
4. Indeling van de ruimte voor spel
5. Ruimte voor privacy
6. Kindergerichte aankleding van de ruimte
7. Ruimte voor actief lichamenlijk spel (groeve motoriek)
8. De speeltrusting voor actief lichamenlijk spel

#### Dagelijkse Individuele verzorging

9. Brengen en halen
10. Maaltijden en tussendoortjes
11. Slapen/rusten
12. Verschonen/toiletgebruik
13. Gezondheidspraktijk
14. Veiligheidspraktijk

#### Taal-Redeneren

15. Boeken en platen
16. Aanmoedigen van de communicatie van kinderen
17. Taal gebruiken voor het ontwikkelen van logisch redeneren
18. Informeel taalgebruik

#### Activiteiten

19. Fijne motoriek
20. Knutselen
21. Muziek en beweging
22. Blokken
23. Zand/water
24. Rollenspel
25. Natuur/wetenschap
26. Rekenen/cijfers
27. Gebruik van tv/video en/of computers
28. Bevorderen van het accepteren van verscheidenheid

#### Interacties

29. Toezicht op actief lichamenlijke activiteiten (groeve motoriek)
30. Algemeen toezicht op kinderen (anders dan groeve motoriek)
31. Discipline (handhaven van orde)
32. Leidster-kind interacties
33. Interacties tussen kinderen

#### Programma

34. Dagschema/dagindeling
35. Vrij spel
36. Groepstijd
37. Voorzieningen voor kinderen met een handicap

#### Ouders en Staf

38. Voorzieningen voor ouders
39. Voorzieningen voor de persoonlijke behoeften van de staf
40. Voorzieningen voor de professionele behoeften van de staf
41. Samenwerking en interactie tussen leidsters
42. Toezicht op en evaluatie van de staf
43. Mogelijkheden voor professionele ontwikkeling



### BIJLAGE 3: Vergelijking van de verschillende items en subschalen

Items	ITERS	ECERS	1995	ITERS-R	ECERS-R
Binnenruimte	-	-		RM	RM
Meubilering voor verzorging, spel en ontwikkeling	RM	RM	RM	RM	RM
Meubilering voor ontspanning en gerief	RM	RM	RM	RM	RM
Indeling van de ruimte voor spel	RM	RM	RM	RM	RM
Ruimte voor privacy	-	SO	I	-	RM
Kindgerichte aankleding van de ruimte	RM	RM	RM	RM	RM
Ruimte voor actief lichamelijk spel	-	FGM	A	-	RM
De speeluitrusting voor actief lichamelijk spel	-	FGM	A	-	RM
Brengen en halen	IV	IV	IV	IV	IV
Maaltijden en tussendoortjes	IV	IV	IV	IV	IV
Slapen/rusten	IV	IV	IV	IV	IV
Verschonen/toiletgebruik	IV	IV	IV	IV	IV
Gezondheidspraktijk	IV	-	IV	IV	IV
Veiligheidspraktijk	IV	-	IV	IV	IV
Persoonlijke verzorging	IV	IV	IV	-	-
Gezondheidsbeleid	IV	-	IV	-	-
Veiligheidsbeleid	IV	-	IV	-	-
Boeken en platen	T	-	T	T	T
Aanmoedigen van de communicatie van kinderen	-	-	-	-	T
Taal gebruiken voor het ontwikkelen van logisch redeneren	-	T	T	-	T
Informeel taalgebruik	T	T	T	T	T
Begrijpen van taal	-	T	T	T	-
Redeneren	-	T	T	-	-
Fijne motoriek	A	FGM	A	A	A
Knutselen	A	CA	A	A	A
Muziek en beweging	A	CA	A	A	A
Blokken	A	CA	A	A	A
Zand/water	A	CA	A	A	A
Rollenspel	A	CA	A	A	A
Natuur/wetenschap	-	-	-	A	A
Rekenen/cijfers	-	-	-	-	A
Gebruik van TV, video, en/of computers	-	-	-	A	A
Bevorderen van het accepteren van verscheidenheid	A	SO	A	A	A
Actief lichamelijk spel	A	-	A	A	
Toezicht op activiteiten voor fijne motoriek	-	FGM	A	-	-
Grove motoriek	-	FGM	A	-	-
Dagschema (creatieve activiteiten)	-	CA	-	-	
Toezicht op creatieve activiteiten	-	CA	P	-	-
Toezicht op actief lichamelijke activiteiten	-	FGM	A	-	I
Algemeen toezicht op kinderen	P	-	P	I	I
Discipline	I	-	I	I	I
Leidster-kind interacties	I	-	I	I	I
Interacties tussen kinderen	I	-	I	I	I
Sfeer	-	SO	I	-	-
Dagindeling/dagschema	P	-	P	P	P
Vrij spel	-	SO	I	P	P
Groepstijd	-	SO	I	P	P
Voorzieningen voor kinderen met een handicap	P	SO	P	P	P
Samenwerking tussen leidsters	P	-	-	OS	OS
Voorzieningen voor ouders	OS	OS	OS	OS	OS
Voorzieningen voor de persoonlijke behoeften van de staf	OS	OS	OS	OS	OS
Voorzieningen voor de professionele behoeften van de staf	-	-	OS	OS	OS
Samenwerking en interacties tussen leidsters	P	-	P	OS	OS
Toezicht op en evaluatie van leidsters	-	-	-	OS	OS
Mogelijkheden voor professionele ontwikkeling	OS	OS	OS	OS	OS
Gemeenschappelijke ruimte voor volwassenen	OS	OS	OS	-	-
Staf continuïteit	-	-	-	OS	-

A = Activiteiten

CA = Creatieve Activiteiten

FGM = Fijne/Grove Motoriek

I = Interacties

IV = Individuele Verzorging

OS = Ouders en Staf

P = Programma

RM = Ruimte en Meubilering

SO = Sociale Ontwikkeling

T = Taal

